

GRUPO II – CLASSE V – Plenário

TC 030.760/2015-1

Natureza: Acompanhamento

Órgão: Ministério do Desenvolvimento Social e Combate à Fome

Representação legal: não há

SUMÁRIO: ACOMPANHAMENTO. CADASTRO ÚNICO PARA PROGRAMAS SOCIAIS (CADÚNICO). PROGRAMA BOLSA FAMÍLIA. VERIFICAÇÃO DA ELEGIBILIDADE DE BENEFICIÁRIOS. AUDITORIA DE DADOS. VERIFICAÇÃO DA CREDIBILIDADE DO CADÚNICO. OPORTUNIDADES DE APRIMORAMENTO. DETERMINAÇÕES. PLANO DE AÇÃO. CIÊNCIA. AUTORIZAÇÃO PARA NOVO ACOMPANHAMENTO. ARQUIVAMENTO.

### Relatório

Trata-se de acompanhamento na concessão, manutenção e pagamento de benefícios assistenciais geridos pela Secretaria Nacional de Renda de Cidadania (Senarc), vinculada ao Ministério do Desenvolvimento Social e Combate à Fome (MDS). Este trabalho foi por mim autorizado nos autos do TC 011.034/2015-7 e realizado pela Secretaria de Controle Externo da Previdência, do Trabalho e da Assistência Social (SecexPrevidência).

2. Acompanhamento é o instrumento utilizado pelo Tribunal para, entre outros objetivos, avaliar, ao longo de um período predeterminado, o desempenho dos de sistemas, programas, projetos e atividades governamentais. Em essência, tem os mesmos atributos de uma auditoria e, no presente caso e em outros similares, vem recebendo a designação de fiscalização (ou auditoria) contínua.

3. Os procedimentos de auditoria foram aplicados no Cadastro Único para Programas Sociais (CadÚnico), que armazena as informações das famílias e das pessoas de baixa renda. O CadÚnico é o instrumento de identificação e de caracterização socioeconômica das famílias brasileiras de baixa renda e é obrigatoriamente utilizado para a seleção de beneficiários de programas sociais do Governo Federal, como por exemplo: Minha Casa Minha Vida; Tarifa Social de Energia Elétrica; e Bolsa Família.

4. Detalhes sobre histórico e lógica da fiscalização contínua constam na íntegra do relatório da SecexPrevidência, acostado à peça 15. Para facilitar a leitura do presente relatório, destaco as etapas do método de análise empregado na auditoria.

5. A fiscalização contínua pode ser dividida nas fases de "obtenção de dados"; "extração, transformação e carga", designada por ETL, sua sigla em inglês; e "análise das informações coletadas", dentro da qual foram desenvolvidas as atividades de credibilidade dos dados, construção dos indicadores gerenciais e tipologias. Antes do teste de incidência nas tipologias, foi realizado um trabalho de *Data Quality* na base do CadÚnico, conhecido na literatura como *Record Linkage*.

6. A verificação de credibilidade permite à equipe de auditoria avaliar se os dados e informações que servem de insumos às análises podem ser utilizados com segurança, sobretudo no sentido de evitar a indevida incidência nas tipologias, caracterizando "falsos positivos".

7. Na transcrição que se segue, com ajustes de forma, enfatizarei os conceitos e relatos necessários à compreensão dos resultados obtidos com a aplicação dos procedimentos de auditoria no CadÚnico.

"(...)

## 2. CADASTRO ÚNICO

### 2.1. VISÃO GERAL

82. O Cadastro Único para Programas Sociais (CadÚnico), instituído pelo Decreto 3.877/2001, que foi revogado e substituído pelo Decreto 6.135/2007, é o instrumento de identificação e de caracterização socioeconômica das famílias brasileiras de baixa renda. O CadÚnico é obrigatoriamente utilizado para a seleção de beneficiários de programas sociais do Governo Federal, como por exemplo: Minha Casa, Minha Vida; Tarifa Social de Energia Elétrica; e Bolsa Família.

83. O Decreto 6.135/2007 define família como a unidade nuclear composta por um ou mais indivíduos, eventualmente ampliada por outros indivíduos que contribuam para o rendimento ou tenham suas despesas atendidas por aquela unidade familiar, todos moradores em um mesmo domicílio.

84. Podem ser incluídas no CadÚnico as famílias que possuem:

a) renda familiar mensal de até meio salário mínimo per capita;

b) renda total de até três salários mínimos;

c) renda superior às alíneas a) ou b), desde que sua inclusão esteja vinculada à participação em programas sociais implementados por quaisquer dos três entes da Federação. Essa permissão busca atender a famílias cuja situação de vulnerabilidade não esteja necessariamente vinculada à renda.

85. O CadÚnico abrange informações, tais como: composição familiar, caracterização do domicílio, identificação e documentação de cada membro da família, escolaridade, trabalho e renda. Essas informações devem ser atualizadas ou revalidadas a cada dois anos.

86. A gestão do Cadastro Único tem por princípio o compartilhamento de esforços e responsabilidades entre a União, os estados, o Distrito Federal e os municípios. Esse modelo tem como base a cooperação e a parceria entre as três esferas de governo, que atuam para fortalecer e consolidar o Cadastro Único como principal instrumento de enfrentamento da pobreza e das desigualdades sociais.

87. No âmbito do Governo Federal, a Secretaria Nacional de Renda e de Cidadania (Senarc) do Ministério do Desenvolvimento Social e Combate à Fome (MDS) acompanha e supervisiona a implantação e a execução do Cadastro Único, realiza a avaliação contínua da qualidade das informações do cadastro e define estratégia para aperfeiçoamento. Também elabora normativos, regulamentos e instruções para orientar o trabalho dos estados, municípios e Distrito Federal.

88. Ainda na esfera federal, a Caixa Econômica Federal (CAIXA), na qualidade de Agente Operador do Programa Bolsa Família, é responsável pelo desenvolvimento do Sistema do Cadastro Único (Cadun), processamento das informações cadastrais, atribuição do Número de Identificação Social (NIS) para cada pessoa cadastrada, com vistas a garantir a unicidade e a integração do cadastro em todos os programas de transferência de renda e assegurar racionalização do processo de cadastramento pelos diversos órgãos públicos.

89. A CAIXA é responsável ainda por disponibilizar o Sistema de Benefícios ao Cidadão (Sibec), com acesso online, oferecer capacitação aos operadores municipais para utilização do sistema e realizar a atualização mensal das informações no Sibec, refletindo as alterações cadastrais processadas no Cadastro Único, a fim de garantir a atualização dos dados e a adequada composição dos benefícios.

90. Outras atividades de responsabilidade da CAIXA são:

a) elaborar, em conjunto com a Senarc, o Calendário Operacional, que apresenta datas e rotinas de processamento essenciais ao funcionamento do CadÚnico e do PBF;

b) atender às famílias beneficiárias via central 0800 e aos gestores municipais;

91. Todos esses processos são planejados, desenvolvidos e executados sob a supervisão e orientação do MDS.

92. Os municípios são os principais atores na gestão do Cadastro Único. A gestão municipal identifica as famílias de baixa renda, efetua o cadastramento, registra os dados na base nacional do Cadastro Único, mantém as informações atualizadas, analisa possíveis inconsistências cadastrais, entre outras atribuições. Os gestores municipais do programa são os principais responsáveis por garantir a realização das ações. Desse modo, organizam e coordenam toda equipe envolvida nas atividades.

93. Os estados atuam no apoio técnico aos municípios relativo à gestão do Cadastro Único.

## 2.2. CARACTERÍSTICAS E ATRIBUIÇÕES

94. O CadÚnico é o instrumento que possibilita a identificação e a caracterização socioeconômica das famílias brasileiras de baixa renda e deve ser obrigatoriamente utilizado para a seleção de beneficiários e para a integração de programas sociais do Governo Federal voltados ao atendimento desse público. A obrigatoriedade de utilização do cadastro não se aplica às políticas contributivas, como o caso da Previdência Social, mas se aplica às políticas não-contributivas e seletivas.

95. Atualmente, o CadÚnico é considerado porta de entrada para mais de vinte programas e serviços sociais em nível federal. Entre esses programas podem-se citar: Programa Bolsa Família (PBF), Benefício de Prestação Continuada (BPC); Serviços de Proteção Social Básica e Especial, Minha Casa Minha Vida (MCMV), Bolsa Verde, Cisternas, Tarifa Social de Energia Elétrica (TSEE) e Fomento às Atividades Produtivas Rurais. Sendo o PBF, o principal usuário das informações do cadastro.

96. O Cadastro Único é composto por:

a) formulários de cadastramento, nos quais as informações das famílias de baixa renda são coletadas pelos entrevistadores municipais;

b) sistema informatizado para a inclusão e a atualização das informações das famílias cadastradas; e

c) base de dados que contém as informações de todas as famílias registradas no CadÚnico em cada município.

97. Esses elementos são fundamentais para que o cadastro possa cumprir sua principal missão: ser um mapa de identificação da parcela mais pobre e vulnerável da população brasileira, trazendo informações para os diferentes territórios sobre suas principais características socioeconômicas, necessidades e potencialidades.

98. Deve-se destacar que o CadÚnico não é um sistema de gestão dos programas ou serviços sociais e nem mesmo um sistema de gestão integrada. Cada órgão gestor, dos diferentes programas sociais, deve constituir equipe para gerir os programas e serviços sociais, utilizando-se das informações presentes no CadÚnico, cujo principal objetivo é identificar e caracterizar as famílias para fornecer uma base de dados que garanta a unicidade das informações para a seleção dos beneficiários de programas sociais.

99. Os fundamentos do CadÚnico são:

a) Instrumento de caracterização socioeconômica para seleção do público beneficiário de programas sociais: o formulário ou a rede de atendimento municipal não se destinam à gestão de benefício. Cada programa social deve possuir a sua própria estrutura para atendimento e gestão dos respectivos benefícios;

b) Padronização dos conceitos: a maioria dos conceitos usados são iguais aos do IBGE, como: família, raça/cor, escolaridade, trabalho;

c) Economicidade do formulário: não possui finalidade de pesquisa. As informações coletadas são utilizadas para fins de elaboração e implementação de políticas públicas para pessoas em situação vulnerável. O formulário deve conter apenas a informação necessária e suficiente para a caracterização socioeconômica;

d) Autodeclaração: a responsabilidade pela veracidade das informações é do responsável familiar (maior de 16 anos). Este fundamento objetiva a quebra da relação de tutela do Estado, a partir da construção de uma relação de confiança com os beneficiários. Ademais, busca-se inibir desvios dos agentes municipais; responsabilizar o cidadão que poderá ser obrigado a devolver os benefícios; e reduzir os custos de verificação e as barreiras à entrada no cadastro;

e) Registro somente com entrevista: garante que a temporalidade do registro esteja circunscrita ao momento da entrevista. Dessa maneira, os dados divergentes de outros registros administrativos são apenas indícios para novas entrevistas.

f) Atualização: necessária frente a qualquer alteração das informações prestadas ou confirmação a cada dois anos.

#### 2.2.1. Cadastramento das famílias

100. O cadastramento das famílias é uma atribuição municipal que requer a participação e o esforço de todos os entes federados para garantir a qualidade das informações coletadas. As atividades de gestão municipal podem ser divididas em quatro etapas:

- a) identificação e localização das famílias vulneráveis;
- b) entrevista e coleta de dados das famílias identificadas;
- c) inclusão dos dados no Sistema do Cadastro Único; e
- d) administração das informações do CadÚnico.

#### 2.2.2. Identificação e localização das famílias

101. A etapa de identificação e localização das famílias corresponde ao planejamento da ação de cadastramento. O primeiro passo é realizar o levantamento preliminar das famílias com base na estimativa de famílias de baixa renda para cada município, disponibilizada pelo MDS aos municípios. A estimativa atual foi calculada com base no Censo de 2010, e a Senarc estuda a possibilidade de realizar atualização dessa estimativa a partir dos dados da Pesquisa Nacional por Amostra Domiciliar (PNAD).

102. O levantamento preliminar tem o objetivo de proporcionar uma referência de quantas famílias com renda até meio salário mínimo moram no município e precisam ser localizadas. A partir dessa referência, o gestor municipal estipula o número de famílias que devem ser entrevistadas, o número de entrevistadores e a infraestrutura necessária.

103. Nessa etapa, também, podem ser usadas outras fontes de dados, como as provenientes das Secretarias Municipais ou Estaduais de Planejamento de Obras; Educação; Saúde ou Assistência Social.

#### 2.2.3. Entrevista e coleta de dados das famílias identificadas

104. O trabalho desenvolvido nessa etapa afeta diretamente a credibilidade das informações da base de dados do CadÚnico. Nos municípios, as responsabilidades por cada etapa do trabalho podem ser divididas entre três atores. O Gestor Municipal do CadÚnico é responsável por coordenar e supervisionar o trabalho das equipes, formadas por entrevistadores e operadores do Sistema do CadÚnico. Os entrevistadores são responsáveis por coletar as informações nos formulários, e os operadores, pela inclusão dessas informações no sistema. Em municípios menores, devido aos custos de contratação de pessoal e ao baixo número de atendimentos realizados, as responsabilidades podem estar concentradas em uma única pessoa.

105. A coleta de dados deve ser realizada de acordo com a metodologia proposta pelo MDS e, para divulgar a metodologia e garantir a qualidade da coleta, a Senarc realiza ações de

capacitação presencial ou à distância para gestores e entrevistadores, além de disponibilizar vídeos, manuais e guias no site oficial do MDS. A capacitação dos operadores, por sua vez, é de responsabilidade da CAIXA.

106. A autodeclaração das informações prestadas pelas famílias constitui um dos fundamentos do CadÚnico. A desvantagem da autodeclaração consiste na dificuldade de comprovar a veracidade das informações. Entretanto, caso fosse exigida uma comprovação mais rigorosa, o número de famílias com acesso ao CadÚnico seria reduzido, uma vez que as famílias cadastradas são de baixa renda e a maioria não participa do mercado formal. O controle do risco da autodeclaração ocorre por meio da atividade de Administração das Informações do CadÚnico, que inclui ações de averiguação, revisão cadastral e exclusão lógica (detalhadas no tópico 2.2.5).

107. A Portaria GM/MDS 177/2011, que trata dos procedimentos para a gestão do CadÚnico, dispõe que a coleta de dados pode ser realizada pelo preenchimento dos formulários ou eletronicamente, com preenchimento direto no Sistema do CadÚnico. No caso de preenchimento dos formulários, deve ser coletada a assinatura do responsável familiar (RF) e do entrevistador após o término da entrevista. No caso de preenchimento eletrônico, é necessário realizar a impressão dos formulários preenchidos ou da folha resumo (anexo 1 da referida portaria) e solicitar a assinatura do RF, do entrevistador e do responsável pelo cadastramento. Os formulários e as folhas resumo devem ser arquivados por um período mínimo de cinco anos em forma física ou eletrônica, após sua digitalização.

108. Os entrevistadores são orientados a respeitar as respostas fornecidas pelo RF e, ao iniciar a entrevista, devem orientar o RF sobre sua responsabilidade em dizer a verdade e sobre o risco de perder o direito de acessar programas sociais e de responder civil ou criminalmente pela omissão de informações ou prestação de informações falsas. Ao final da entrevista, o RF é informado que a inscrição no CadÚnico não garante a inclusão automática em nenhum programa que faz uso das informações do cadastro.

109. Em relação às equipes municipais, a Senarc informou que possuem alta rotatividade e, por esse motivo, a secretaria não dispõe de uma estimativa confiável do número de entrevistadores ou operadores capacitados em atividade.

#### Formas de Coletas de Dados

110. A coleta de dados é realizada das seguintes formas:

- a) visitas domiciliares;
- b) postos de atendimento fixos;
- c) postos de atendimento itinerantes.

111. A visita domiciliar é a forma mais indicada para realizar o cadastramento das famílias de baixa renda, pois permite observar o ambiente familiar no momento de realizar o preenchimento do formulário. O percentual de formulários preenchidos por visita familiar é de 6%, considerando o total de famílias cadastradas. O art. 5º, §2º da Portaria GM/MDS 177/2011, entretanto, determina que o número de famílias cadastradas por meio de visita domiciliar deve corresponder a pelo menos 20% do total.

112. Os postos de atendimento fixos são locais disponibilizados pelos municípios para que as famílias se inscrevam no CadÚnico ou atualizem seus dados. Um município pode possuir mais de um posto de atendimento fixo, usualmente localizados na sede da prefeitura, no Centro de Referência de Assistência Social (CRAS) ou em estrutura própria.

113. Os postos de atendimento itinerantes são unidades móveis para atendimento do público por período determinado, que percorrem regiões distantes e áreas de difícil acesso. O cadastramento por meio de Mutirão está incluído na categoria de postos itinerantes e tem o objetivo de realizar ações pontuais de cadastramento, como, por exemplo, de assentados ou de catadores de material reciclável.

### Documentos necessários para o cadastramento

114. O RF deve apresentar, obrigatoriamente, o CPF ou o título de eleitor para que a família seja cadastrada. Os demais componentes da família devem apresentar, no mínimo, um dos seguintes documentos: Certidão de Nascimento, Certidão de Casamento, CPF, RG, carteira de trabalho, título de eleitor ou Certidão Administrativa de Nascimento Indígena. Apesar da exigência de apenas um documento, os entrevistadores são instruídos a coletar as informações de todos os documentos apresentados.

115. Caso o RF ou demais componentes não possuam os documentos necessários para o cadastramento, o entrevistador pode realizar a coleta das demais informações e comunicar que o cadastro será marcado como 'sem registro civil'. Nesse caso, o cadastrado tem o prazo de até 30 dias para apresentar a documentação, o que permitirá sua participação em programas sociais que fazem uso do CadÚnico.

116. Se os integrantes da família não possuírem documentação por falta de registro civil, eles devem ser encaminhados para a realização do registro. Caso já possuam o registro civil, devem ser encaminhados para expedição da 2ª via da documentação necessária ao cadastramento.

117. Para as famílias indígenas ou quilombolas, o RF pode apresentar, no mínimo, um dos documentos requeridos aos demais componentes da família.

#### 2.2.4. Inclusão de dados no Sistema do CadÚnico

118. O processo de inclusão dos dados cadastrais no sistema é composto pela digitação dos formulários no sistema, pela atribuição de código familiar e pela atribuição e validação do Número de Identificação Social (NIS) para cada pessoa.

119. A digitação dos formulários é realizada pelos operadores do sistema e, em alguns casos, um grupo de municípios envia os formulários para um município com melhor estrutura, que fica responsável pela digitação.

120. O código familiar é atribuído pelo sistema no momento da inclusão de uma nova família.

121. O processo de atribuição e validação do NIS é realizado pela CAIXA, conforme as regras de unicidade previamente definidas, por meio do Sistema do CadÚnico e dos sistemas corporativos.

122. O Sistema do CadÚnico não é um sistema independente; quando foi desenvolvido pela CAIXA, optou-se por criá-lo como um módulo dos sistemas corporativos existentes. Como exemplos de sistemas existentes, pode-se citar o CADNIS, responsável pela gestão do NIS, e os sistemas do Programa de Integração Social (PIS), Programa de Formação do Patrimônio do Servidor Público (PASEP) e carteira de trabalho.

123. O primeiro passo para a atribuição do NIS a cada pessoa cadastrada é a geração de um número provisório, que será confrontado com os NIS existentes no CADNIS, com o objetivo de garantir a unicidade do registro. Caso o cruzamento de informações detecte duplicidade de registros, será atribuído o NIS já existente ao registro. Caso não ocorra duplicidade, o NIS provisório é validado. Ao final desse processo, o registro da pessoa é marcado como 'cadastrado'.

#### 2.2.5. Administração das informações do CadÚnico

124. A administração das informações cadastrais é importante para garantir a qualidade dos dados e assegurar que as informações registradas na base estejam de acordo com a realidade das famílias. A atuação da Senarc nessa etapa pode ser dividida em três ações: Revisão Cadastral, Averiguação Cadastral e Exclusão Lógica por desatualização.

125. A Revisão Cadastral é uma ação prevista no art. 7º do Decreto 6.135/2007 e regulamentada pela Portaria GM/MDS 617/2010. Por meio dessa ação, as famílias devem

atualizar ou revalidar as informações sempre que houver alteração em campos específicos do seu cadastro (composição familiar, endereço, renda, documentação do RF, mudança de escola) ou no prazo máximo de 24 meses, contados da data da última entrevista.

126. No Sistema do CadÚnico existem os campos de atualização cadastral e de alteração cadastral, porém ambos não se confundem. O primeiro, registra a data da última atualização ou revalidação das informações. O segundo não está relacionado com esses parâmetros, sendo de uso interno da CAIXA.

127. A Averiguação Cadastral é regulamentada pela Portaria GM/MDS 94/2013 e tem o objetivo de verificar a consistência das informações cadastrais por meio da análise de dados e do cruzamento com outras bases, bem como de desencadear medidas para o tratamento das inconsistências. Além das inconsistências detectadas pela Senarc, a averiguação inclui as inconsistências que são apontadas pelos órgãos de controle ou entidades públicas detentoras de bases de dados, cujos registros possam ser comparados aos do CadÚnico.

128. A Exclusão Lógica por desatualização é a marcação do registro de uma família como 'excluído' no CadÚnico, caso não ocorra a atualização ou a revalidação das informações cadastrais no prazo de 48 meses. É importante destacar a diferença entre a exclusão lógica e a exclusão física do registro. No primeiro caso, o registro é marcado como 'excluído' e a família permanece na base, mas não pode participar dos programas até nova atualização. No segundo caso, o registro da família é excluído da base.

### 2.3. BASE DE DADOS DO CADASTRO ÚNICO

#### 2.3.1. Obtenção dos dados e processo de ETL [extração, transformação e carga]

129. As bases de dados do CadÚnico e da Folha de pagamentos do PBF foram solicitadas à Senarc (...) no âmbito do processo de Produção de Conhecimento TC 011.598/2015-8.

130. As informações foram enviadas a partir de um container criptografado com o *software* Truecrypt e, posteriormente, a sua integridade foi verificada pelo número de hash [sequência de números que assegura que a informação recebida foi aquela enviada pelo remetente sem adulteração].

131. A SecexPrevidência foi auxiliada pelo Sepin/Digep, que ficou responsável pelo recebimento dos arquivos, a verificação do hash e o processo de ETL para a disponibilização das informações no servidor SQL do Labcontas.

#### 2.3.2. Estrutura da Base de Dados do CadÚnico

132. A base de dados do CadÚnico foi desenvolvida para armazenar as informações coletadas pelo formulário de cadastramento. Os formulários são divididos em formulário principal e dois formulários suplementares. O formulário principal é dividido em 10 blocos, conforme a Tabela 01.

Tabela 01. Blocos do Formulário Principal de Cadastramento

Bloco 1	Identificação e Controle	
Bloco 2	Características do domicílio	Identificação da família e do domicílio
Bloco 3	Família	
Bloco 4	Identificação da pessoa	Identificação da pessoa
Bloco 5	Documentos	
Bloco 6	Pessoas com deficiência	
Bloco 7	Escolaridade	
Bloco 8	Trabalho e Remuneração	
Bloco 9	Responsável pela unidade familiar – RF	Assinatura do RF
Bloco 10	Marcação livre para o município	Identificação de trabalho infantil

133. Os formulários suplementares têm o objetivo de coletar informações sobre a vinculação das famílias a programas e serviços e sobre as pessoas em situação de rua.

134. Para armazenar as informações, a base de dados do CadÚnico foi estruturada em dezessete tabelas que podem ser divididas em quatro partes.

a) Tabelas referentes as famílias. Composta por três tabelas principais que armazenam as informações dos Blocos 1, 2 e 3. Sendo que cada tabela possui 36,3 milhões de registros.

b) Tabela referentes as Pessoas. Composta por cinco tabelas principais que armazenam as informações dos Blocos 4 ao 8. Cada tabela possui aproximadamente 133,1 milhões de registros.

c) Tabela de Elos. Tabela que armazena as informações sobre a mudança de NIS das pessoas cadastradas e possui 13,3 milhões de registros.

d) Tabelas com as informações dos formulários suplementares e outras tabelas com informações sobre pendências de preenchimento, transferência de famílias ou pessoas e demais informações.

### 3. EXAME TÉCNICO

135. As análises realizadas na base de dados do CadÚnico estão relacionadas com o conceito de *Data Quality*, que se refere ao nível de qualidade dos dados. Pode-se afirmar que os dados têm alta qualidade se estão 'aptos para o uso' nas suas necessidades operacionais, de planejamento ou para tomada de decisão (Juran and Godfrey, Juran's Quality Handbook, 1999).

136. O escopo do trabalho foi a base de dados do CadÚnico, que é o instrumento para a seleção de beneficiários em mais de vinte programas sociais, sendo os principais: Programa Bolsa Família, Minha Casa, Minha Vida e Tarifa Social de Energia Elétrica.

137. Ao avaliar e contribuir para o aumento da qualidade dos dados, objetiva-se a melhoria no processo de seleção dos beneficiários, bem como, o aumento da efetividade dos programas sociais ao atingir as famílias e indivíduos em situação de fragilidade social.

#### 3.1. ANÁLISE DE CREDIBILIDADE DAS INFORMAÇÕES DO CADASTRO ÚNICO

##### 3.1.1. Objetivo

138. As informações do banco de dados do CadÚnico serão avaliadas quanto à sua credibilidade, que é composta por seis atributos: completude, unicidade, validade, consistência, acurácia e uniformidade.

139. A completude é a verificação da existência de registros com dados faltantes. Há casos em que essa falta é justificável, por exemplo, o campo 'CPF' não precisa estar preenchido para todos os membros da família.

140. A unicidade do campo verifica se existe duplicidade na chave primária da tabela (simples ou composta). Alguns exemplos são o número de CPF ou CNPJ na respectiva base da Receita Federal e a chave composta por Código de Prefeitura, Código de Família e Código de Pessoa no Cadastro Único.

141. A validade busca identificar se o esquema do banco de dados está sendo respeitado. Verificam se campos que deveriam ser numéricos estão sendo disponibilizados dessa forma; se números de CPF e título de eleitor respeitam a regra 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.

142. 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 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 (um número de NIS sendo utilizado por duas pessoas diferentes, por exemplo).

143. A acurácia dos dados é verificada por meio de testes que confrontem os dados obtidos com outras fontes de informação. CPFs e informações cadastrais ('nome', 'nome da mãe' e 'data de nascimento') do CadÚnico podem ser comparados com seus equivalentes nos bancos de dados da Receita Federal.

144. A uniformidade verifica se campos com informações quantitativas mantêm um padrão de medida. Esses testes podem ser feitos a campos numéricos monetários (descrições em Reais, Centavos, Dólares, Cruzeiros) ou baseados em outras unidades (quilo, litro, resma, etc.), conforme o caso.

### 3.1.2. Dados Utilizados

145. O trabalho utilizou a base do Cadastro Único referente ao mês de maio de 2015, que foi enviada pela Senarc (...).

146. Conforme instruções do MDS (peça 08), somente foram considerados os registros das famílias e as pessoas com estado cadastral igual a 'cadastrado' (...).

147. A base de dados utilizada possui o registro de 27,3 milhões de famílias e 82 milhões de pessoas, conforme indicam as Tabelas 02 e 03, que apresentam a quantidade famílias e pessoas por tipo de código cadastral. Os registros marcados como 'Em cadastramento' ainda não finalizaram todas as etapas de cadastramento. Caso os documentos obrigatórios para finalizar o cadastro não sejam apresentados, o registro é marcado como 'Sem registro civil'. Os registros em 'aguardando NIS' e 'validando NIS' estão na fase de atribuição do NIS e do confronto das informações declaradas, com as informações existentes no sistema do cadastro NIS (CADNIS).

Tabela 02. Famílias por tipo de código cadastral

Estado Cadastral	Qtd.
1 - Em cadastramento	25.800
2 - Sem registro civil	2.718
3 - Cadastrado	27.378.606
4 - Excluído	8.952.092
Total	36.359.216

Tabela 03. Pessoas por tipo de código cadastral

Estado Cadastral	Qtd.
1 - Em Cadastramento	39.519
2 - Sem registro civil	32.871
3 - Cadastrado	82.037.928
4 - Excluído	50.867.061
5 - Aguardando NIS	45.265
6 - Validando NIS	129.414
Total	133.152.058

### 3.1.3. Verificação da unicidade

148. No CadÚnico, os campos de Código Prefeitura, Código Família e Código Pessoa são chaves primárias na tabela e a união dos três campos forma uma chave composta. A relação entre os campos pode ser entendida como subconjuntos, as pessoas estão contidas em famílias, que por sua vez estão contidas em prefeituras.

149. Nos testes realizados, os campos apresentaram unicidade como chave primária simples e como chave primária composta.

### 3.1.4. Verificação da completude e de validade

150. Os atributos de completude e de validade foram analisados em conjunto, por meio da verificação das regras de preenchimento do formulário, conforme as instruções disponíveis no Manual do Entrevistador (peça 10) e no Dicionário de Dados v.5.41 (peça 09), enviado pela Senarc junto com a base de dados do CadÚnico.

151. Para cada uma das tabelas que compõem a estrutura do banco de dados foram desenvolvidos scripts em SQL que avaliam o preenchimento dos campos e retornam o valor '1' caso ocorra alguma inconsistência.

#### 3.1.4.1. Bloco 1 - Identificação e Controle

152. O Bloco 1 possui informações sobre a forma de preenchimento do formulário, endereço da família e identificação da pessoa responsável pela entrevista.

153. A Tabela 04 indica que existem inconsistências em 9,76% dos registros familiares em relação ao preenchimento das informações do CEP, campo NUM\_CEP\_LOGRADOURO\_FAM, e inconsistências em 7,57% dos registros em relação à identificação do entrevistador, campos NOM\_ENTREVISTADOR\_FAM e NUM\_CPF\_ENTREVISTADOR\_FAM.

Tabela 04 – Quantidade de inconsistências para o Bloco 1

Bloco	Campo	Quantidade de inconsistências	%
1	DAT_CADASTRAMENTO_FAM	11.066	0,04
1	DAT_ATUAL_FAM	0	0,00
1	COD_EST_CADASTRAL_FAM	0	0,00
1	IND_CADASTRO_VALIDO_FAM	0	0,00
1	COD_CONDICAO_CADASTRO_FAM	0	0,00
1	VLR_RENDA_MEDIA_FAM	0	0,00
1	COD_MUNIC_IBGE_2_FAM	0	0,00
1	COD_MUNIC_IBGE_5_FAM	0	0,00
1	COD_MODALIDADE_FAM	0	0,00
1	COD_FORMA_COLETA_FAM	0	0,00
1	IND_FORMULARIO_0_FAM	0	0,00
1	IND_FORMULARIO_1_FAM	0	0,00
1	IND_FORMULARIO_2_FAM	0	0,00
1	IND_FORMULARIO_SUP1_FAM	0	0,00
1	IND_FORMULARIO_SUP2_FAM	0	0,00
1	DTA_ENTREVISTA_FAM	146	< 0,01
1	NOM_LOCALIDADE_FAM	2.671.745	9,76
1	NOM_TIP_LOGRADOURO_FAM	2.671.752	9,76
1	NOM_LOGRADOURO_FAM	2.671.745	9,76
1	DES_COMPLEMENTO_FAM	2.671.804	9,76
1	NUM_CEP_LOGRADOURO_FAM	2.671.748	9,76
1	NOM_ENTREVISTADOR_FAM	2.071.811	7,57
1	NUM_CPF_ENTREVISTADOR_FAM	2.071.811	7,57

154. A auditoria identificou a ausência do preenchimento do campo CEP (NUM\_CEP\_LOGRADOURO\_FAM) e nos campos referentes ao endereço (NOM\_LOCALIDADE\_FAM, NOM\_TIP\_LOGRADOURO\_FAM, NOM\_LOGRADOURO\_FAM e DES\_COMPLEMENTO\_FAM) em 9,76% dos registros.

#### ACHADO 02 – INCONSISTÊNCIA NO PREENCHIMENTO DO NOME E DO NÚMERO DE CPF DO ENTREVISTADOR

155. A auditoria identificou a ausência do preenchimento do campo referente ao nome (NOM\_ENTREVISTADOR\_FAM) e ao número de CPF do entrevistador (NUM\_CPF\_ENTREVISTADOR\_FAM) em 7,57% dos registros.

##### 3.1.4.2. Bloco 2 - Características do domicílio

156. O Bloco 2 registra informações em relação às características do domicílio da família como a quantidade de cômodos, material de revestimento dos pisos e paredes e o atendimento pelos serviços de água, energia e coleta de lixo.

157. Campo COD\_ESPECIE\_DOMIC\_FAM é de preenchimento obrigatório e identifica a espécie do domicílio entre as opções: particular permanente, particular improvisado ou coletivo. A análise do preenchimento do campo identificou que as inconsistências decorrem do seu não preenchimento, que devido regras de preenchimento do formulário, afeta a análise dos demais campos, que apresentaram em média 8,3% de inconsistências no seu preenchimento, conforme apresenta a Tabela 05

158. Os campos que informam a quantidade de cômodos no domicílio (QTD\_COMODOS\_DOMIC\_FAM) e quantidade de cômodos usados como dormitório (QTD\_COMODOS\_DORMITORIO\_FAM) apresentam valores absurdos na resposta, como

domicílios com mais de 50 cômodos ou dormitórios, e possuem 11,35% dos registros com inconsistência.

Tabela 05 – Quantidade de inconsistências para o Bloco 2

Bloco	Campo	Quantidade de inconsistências	%
2	COD_LOCAL_DOMIC_FAM	461	< 0,01
2	COD_ESPECIE_DOMIC_FAM	2.106.918	7,70
2	QTD_COMODOS_DOMIC_FAM	3.111.438	11,36
2	QTD_COMODOS_DORMITORIO_FAM	3.108.115	11,35
2	COD_MATERIAL_PISO_FAM	2.106.932	7,70
2	COD_MATERIAL_DOMIC_FAM	2.107.762	7,70
2	COD_AGUA_CANALIZADA_FAM	2.107.735	7,70
2	COD_ABASTE_AGUA_DOMIC_FAM	2.107.727	7,70
2	COD_BANHEIRO_DOMIC_FAM	2.107.668	7,70
2	COD_ESCOA_SANITARIO_DOMIC_FAM	3.165.363	11,56
2	COD_DESTINO_LIXO_DOMIC_FAM	2.107.752	7,70
2	COD_ILUMINACAO_DOMIC_FAM	2.107.718	7,70
2	COD_CALCAMENTO_DOMIC_FAM	2.671.566	9,76

### ACHADO 03 - INCONSISTÊNCIA NO PREENCHIMENTO DO CAMPO COD\_ESPECIE\_DOMIC\_FAM

159. A auditoria identificou a ausência do preenchimento do campo referente à espécie de domicílio (COD\_ESPECIE\_DOMIC\_FAM) em 7,7% dos registros, que afeta o preenchimento dos demais campos do bloco e aumenta o número de inconsistências.

#### 3.1.4.3. Bloco 3 – Família

160. O Bloco 3 registra informações sobre o número e nome dos componentes da família, se a família é indígena ou quilombola, bem como, os valores referentes a despesa mensal da família com energia elétrica, água e esgoto, alimentação, transporte, aluguel e medicamentos.

161. Os itens em negrito na Tabela 06 estão com um percentual elevado de inconsistências. Apesar das informações estarem preenchidas nos campos, a análise das regras de relacionamento entre os campos indica que ocorreram problemas na padronização do preenchimento em relação as regras de preenchimento indicadas no Dicionário de Dados v.5.41 (peça 09).

Tabela 06 – Quantidade de inconsistências para o Bloco 3

Bloco	Campo	Quantidade de inconsistências	%
3	COD_FAMILIA_INDIGENA_FAM	80	< 0,01
3	COD_POVO_INDIGENA_FAM	14.079	0,05
3	NOM_POVO_INDIGENA_FAM	80	< 0,01
3	COD_INDIGENA_RESIDE_FAM	14.069	0,05
3	<b>COD_RESERVA_INDIGENA_FAM</b>	<b>27.292.949</b>	<b>99,69</b>
3	<b>NOM_RESERVA_INDIGENA_FAM</b>	<b>27.292.949</b>	<b>99,69</b>
3	<b>IND_RESERVA_INDIGENA_FAM</b>	<b>27.292.949</b>	<b>99,69</b>
3	IND_FAMILIA_QUILOMBOLA_FAM	128.245	0,47
3	<b>COD_COMUNIDADE_QUILOMBOLA_FAM</b>	<b>27.242.646</b>	<b>99,50</b>
3	<b>NOM_COMUNIDADE_QUILOMBOLA_FAM</b>	<b>27.242.646</b>	<b>99,50</b>
3	<b>IND_COMUNIDADE_QUILOMBOLA_FAM</b>	<b>27.242.646</b>	<b>99,50</b>
3	QTD_PESSOAS_DOMIC_FAM	2.722.610	9,94
3	QTD_FAMILIAS_DOMIC_FAM	2.732.768	9,98
3	QTD_PESSOA_INTER_0_17_ANOS_FAM	1.006	< 0,01
3	QTD_PESSOA_INTER_18_64_ANOS_FA	1.085	< 0,01
3	QTD_PESSOA_INTER_65_ANOS_FAM	1.305	< 0,01
3	VAL_DESP_ENERGIA_FAM	404.189	1,48
3	VAL_DESP_AGUA_ESGOTO_FAM	862.295	3,15
3	VAL_DESP_GAS_FAM	314.712	1,15
3	VAL_DESP_ALIMENTACAO_FAM	64.388	0,24
3	VAL_DESP_TRANSPOR_FAM	2.026.602	7,40
3	VAL_DESP_ALUGUEL_FAM	1.950.255	7,12

3 VAL\_DESP\_MEDICAMENTOS\_FAM 1.731.386 6,32

#### ACHADO 04 – INCONSISTÊNCIA NAS REGRAS DE PREENCHIMENTO DOS CAMPOS DO BLOCO 3 [família]

162. Identificou-se que os campos listados na Tabela 07 estão preenchidos. Contudo, ao aplicar as regras de preenchimento do Dicionários de Dados (peça 09), que levam em consideração o preenchimento conjunto dos campos, foram detectadas inconsistências no preenchimento em 99% dos campos.

163. Um exemplo de inconsistência é o preenchimento nome e do código da reserva indígena (COD RESERVA INDIGENA FAM e NOM RESERVA INDIGENA FAM), enquanto a opção 'não sabe informar a reserva' (IND RESERVA INDIGENA FAM) está marcada.

Tabela 07 – Campos com inconsistência no registro

Campo	Quantidade de inconsistências	de Qtd. de famílias (%)	Qtd. de inconsistências / Total (%)
COD_RESERVA_INDIGENA_FAM	27.292.949		99,69
NOM_RESERVA_INDIGENA_FAM	27.292.949		99,69
IND_RESERVA_INDIGENA_FAM	27.292.949		99,69
COD_COMUNIDADE_QUILOMBOLA_FAM	27.242.646		99,50
NOM_COMUNIDADE_QUILOMBOLA_FAM	27.242.646		99,50
IND_COMUNIDADE_QUILOMBOLA_FAM	27.242.646		99,50

#### 3.1.4.4. Bloco 4 - Identificação da pessoa

164. O Bloco 4 registra informações sobre a pessoa, como o nome, a data de nascimento, o sexo, o nome da mãe e do pai, a raça e o Número de Identificação Social (NIS).

165. Os campos do bloco apresentam um percentual baixo de inconsistências. Contudo, os campos marcados em negrito na Tabela 08, estão com percentual elevado de inconsistências, quando comparados com as regras de preenchimento indicadas no Dicionário de Dados v.5.41 (peça 09).

Tabela 08 – Quantidade de inconsistências para o Bloco 4

Bloco	Campo	Quantidade de inconsistências	de %
4	DTA_CADASTRAMENTO_MEMB	1.851	< 0,01
4	DTA_ATUAL_MEMB	1.334	< 0,01
4	COD_EST_CADASTRAL_MEMB	0	0,00
4	NUM_ORDEM_PESSOA	3	< 0,01
4	NOM_PESSOA	0	0,00
4	NUM_NIS_PESSOA_ATUAL	0	0,00
4	COD_SEXO_PESSOA	0	0,00
4	DTA_NASC_PESSOA	635	< 0,01
4	COD_PARENTESCO_RF_PESSOA	528.629	0,64
4	COD_RACA_COR_PESSOA	845.736	1,03
4	NOM_COMPLETO_MAE_PESSOA	304	< 0,01
4	IND_NOM_COMPLETO_MAE_PESSOA	304	< 0,01
4	NOM_COMPLETO_PAI_PESSOA	491.735	0,60
4	IND_NOM_COMPLETO_PAI_PESSOA	491.735	0,60
4	COD_LOCAL_NASCIMENTO_PESSOA	344.967	0,42
4	<b>SIG_UF_MUNIC_NASC_PESSOA</b>	<b>45.394.568</b>	<b>55,33</b>
4	<b>IND_UF_MUNIC_NASC_PESSOA</b>	<b>45.394.568</b>	<b>55,33</b>
4	<b>NOM_IBGE_MUNIC_NASC_PESSOA</b>	<b>45.394.568</b>	<b>55,33</b>
4	<b>COD_IBGE_MUNIC_NASC_PESSOA</b>	<b>45.394.568</b>	<b>55,33</b>
4	<b>IND_IBGE_MUNIC_NASC_PESSOA</b>	<b>45.394.568</b>	<b>55,33</b>
4	<b>NOM_PAIS_ORIGEM_PESSOA</b>	<b>82.033.634</b>	<b>99,99</b>
4	<b>COD_PAIS_ORIGEM_PESSOA</b>	<b>82.033.634</b>	<b>99,99</b>
4	<b>IND_PAIS_ORIGEM_PESSOA</b>	<b>82.033.634</b>	<b>99,99</b>
4	COD_CERTIDAO_REGISTRADA_PESSOA	4.326	0,01

#### ACHADO 05 – INCONSISTÊNCIA NAS REGRAS DE PREENCHIMENTO DOS CAMPOS DO BLOCO 4 [Identificação da pessoa]

166. A situação encontrada identificou que os campos listados na Tabela 09 estão preenchidos. Contudo ao aplicar as regras de preenchimento do Dicionários de Dados (peça 09), que levam em consideração o preenchimento conjunto dos campos, foram detectadas inconsistências no preenchimento dos campos acima de 55% do total de registros.

167. Um exemplo de inconsistência é o preenchimento do nome e do código de município (NOM\_IBGE\_MUNIC\_NASC\_PESSOA e COD\_IBGE\_MUNIC\_NASC\_PESSOA), enquanto a opção 'não sabe informar o município' (IND\_IBGE\_MUNIC\_NASC\_PESSOA) está marcada.

Tabela 09 – Campos com inconsistência no registro

Campo	Quantidade de inconsistências	Qtd. de inconsistências / Total de famílias (%)
SIG_UF_MUNIC_NASC_PESSOA	45.394.568	55,33
IND_UF_MUNIC_NASC_PESSOA	45.394.568	55,33
NOM_IBGE_MUNIC_NASC_PESSOA	45.394.568	55,33
COD_IBGE_MUNIC_NASC_PESSOA	45.394.568	55,33
IND_IBGE_MUNIC_NASC_PESSOA	45.394.568	55,33
NOM_PAIS_ORIGEM_PESSOA	82.033.634	99,99
COD_PAIS_ORIGEM_PESSOA	82.033.634	99,99
IND_PAIS_ORIGEM_PESSOA	82.033.634	99,99

#### 3.1.4.5. Bloco 5 – Documentos

168. O Bloco 5 registra as informações sobre os documentos da pessoa. O formulário possui campos para o preenchimento de informações sobre as certidões de nascimento, de casamento ou de nascimento do indígena, o número do CPF, o documento de identidade (RG), a carteira de trabalho e os dados do título de eleitor.

169. Para o Responsável Familiar (RF), o preenchimento das informações do CPF ou do Títulos de Eleitor é obrigatório. Para os demais componentes da família e o RF de famílias indígenas ou quilombolas é obrigatório o preenchimento de um dos seguintes documentos: Certidão de Nascimento, Certidão de Casamento, CPF, RG, carteira de trabalho, título de eleitor ou Certidão Administrativa de Nascimento Indígena. A variável REQ\_MIN\_DOCUMENTO, na Tabela 10, indica que 4,46% das pessoas não cumprem os requisitos mínimos para o preenchimento dos documentos.

170. A análise dos campos indica que a maioria das inconsistências decorre de erros no preenchimento dos campos de nome do cartório, número do livro, folha e termo e data do registro das certidões de nascimento ou de casamento.

Tabela 10 – Quantidade de inconsistências para o Bloco 5

Bloco	Campo	Quantidade de inconsistências	%
5	COD_CERTIDAO_CIVIL_PESSOA	0	0,00
5	REQ_MIN_DOCUMENTO	3.661.286	4,46

#### ACHADO 06 – REGISTROS QUE NÃO CUMPREM AS REGRAS MÍNIMAS PARA PREENCHIMENTO DOS DOCUMENTOS

171. O preenchimento dos campos de documentos no formulário do cadastro único possui regras específicas caso o indivíduo seja o Representante Familiar ou pertença aos demais membros da família, bem como, existem regras para famílias indígenas, quilombolas e em situação da rua.

172. O campo REQ\_MIN\_DOCUMENTO identifica as pessoas que não cumpriram os requisitos mínimos de preenchimento, conforme as regras do Dicionário de Dados v5.41 (peça 09), e corresponde a 4,46% do total de registro de indivíduos.

173. A Tabela 11 apresenta a situação encontrada pelo tipo de família, tipo de pessoa e tipo de documentos. Os valores para o tipo de documento informam se o documento foi preenchido. Para o RF da família de baixa renda, foram identificados 240.529 registros com o

preenchimento indevido, nos quais, se observa o preenchimento de outros documentos, contudo as informações de CPF e título de eleitor não estão preenchidas.

174. Para os demais membros das famílias de baixa renda, foram identificados 3,37 milhões de registros com preenchimento indevido e a análise das informações do banco de dados identificou inconsistências no preenchimento dos campos de nome do cartório, número do livro, folha e termo e data do registro para as certidões de nascimento ou de casamento.

Tabela 11 – Quantidade de pessoas que não cumprem as regras de preenchimento dos documentos

Família em situação de rua	Família	Pessoa	Quantidade	Tipo de Documento				
				Certidão	CPF	Cart. Ident	Cart. Trab	Tit. Eleit.
Não	Baixa Renda		240.529	218.536	0	49.625	7.614	0
	Indígena	Resp. Familiar	368	0	0	0	0	0
	Quilombola		49	0	0	0	0	0
	Baixa Renda		3.375.998	0	0	0	0	0
	Indígena	Demais membros	28.190	0	0	0	0	0
	Quilombola		15.934	0	0	0	0	0
Sim	Baixa Renda	Resp. Familiar	68	60	0	11	2	0
	Baixa Renda		148	0	0	0	0	0
	Indígena	Demais membros	2	0	0	0	0	0
Total			3.661.286	218.596	0	49.636	7.616	0

#### 3.1.4.6. Campo CPF ou título de eleitor preenchido com número inválido

175. Os números de CPF e de título de eleitor são gerados a partir de uma regra matemática que permite a sua verificação. O teste permite identificar se o número do documento é possível, mas não permite afirmar se o número existe ou está em situação regular. A equipe de auditoria desenvolveu os algoritmos para verificar a validade dos documentos e, caso o campo não obedeça a regra de verificação, o documento é considerado inválido.

176. A Tabela 12 apresenta o total de pessoas com o número de CPF ou de título de eleitor inválido detectado no CadÚnico. O número de pessoas com documentos inválidos é baixo em relação ao número de documentos preenchidos.

Tabela 12 – Pessoas com o CPF ou o título de eleitor inválido

Tipo de documento	Total de pessoas com o documento inválido	Total de pessoas com o documento preenchido
CPF	50	46.892.630
título de eleitor	4.056	34.073.980

#### ACHADO 07 – PESSOAS COM NÚMERO DE CPF E TÍTULO DE ELEITOR INVÁLIDO.

177. A equipe de auditoria identificou 50 números de CPF e 4056 números de título de eleitor inválidos por não cumprirem a regra de formação.

#### ACHADO 08 – CAMPO DO TÍTULO DE ELEITOR COM NÚMERO DE DÍGITOS SUPERIOR AO DEFINIDO PELO TRIBUNAL SUPERIOR ELEITORAL

178. O campo NUM\_TITULO\_ELEITOR\_PESSOA para o preenchimento das informações do título de eleitor da pessoa é definido como composto por 13 dígitos, conforme o dicionário de dados (peça 09). Contudo o Tribunal de Superior Eleitoral define o número como composto por 12 dígitos (peça 11), essa diferença no registro dificulta a validação da regra de formação do número do título de eleitor e contribui para o registro inadequado da informação.

#### 3.1.4.7. Bloco 6 - Pessoas com deficiência

179. O Bloco 6 registra as informações em caso de deficiência na pessoa cadastrada. Conforme indica a Tabela 13, não foram detectadas inconsistências no bloco.

Tabela 13 – Quantidade de inconsistências para o Bloco 6

Bloco	Campo	Quantidade de inconsistências	%
6	COD_DEFICIENCIA_MEMB	0	0,00
6	IND_DEF_CEGUEIRA_MEMB	0	0,00
6	IND_DEF_BAIXA_VISAO_MEMB	0	0,00
6	IND_DEF_SURDEZ_PROFUNDA_MEMB	0	0,00
6	IND_DEF_SURDEZ_LEVE_MEMB	0	0,00
6	IND_DEF_FISICA_MEMB	0	0,00
6	IND_DEF_MENTAL_MEMB	0	0,00
6	IND_DEF_SINDROME_DOWN_MEMB	0	0,00
6	IND_DEF_TRANSTORNO_MENTAL_MEMB	0	0,00
6	IND_AJUDA_NAO_MEMB	0	0,00
6	IND_AJUDA_FAMILIA_MEMB	0	0,00
6	IND_AJUDA_ESPECIALIZADO_MEMB	0	0,00
6	IND_AJUDA_VIZINHO_MEMB	0	0,00
6	IND_AJUDA_INSTITUICAO_MEMB	0	0,00
6	IND_AJUDA_OUTRA_MEMB	0	0,00

### 3.1.4.8. Bloco 7 – Escolaridade

180. As informações sobre escolaridade e sobre qual escola a pessoa frequenta são registrados no Bloco 7. A tabela 14 indica que ocorre um baixo percentual de inconsistência neste bloco, a média de inconsistências no preenchimento das variáveis é de 1,1%.

Tabela 14 – Quantidade de inconsistências para o Bloco 7

Bloco	Campo	Quantidade de inconsistências	%
7	COD_SABE_LER_ESCREVER_MEMB	506.754	0,62
7	IND_FREQUENTA_ESCOLA_MEMB	831.944	1,01
7	NOM_ESCOLA_MEMB	1.014.147	1,24
7	COD_ESCOLA_LOCAL_MEMB	1.054.388	1,29
7	SIG_UF_ESCOLA_MEMB	1.054.388	1,29
7	NOM_MUNIC_ESCOLA_MEMB	1.054.389	1,29
7	COD_IBGE_MUNIC_ESCOLA_MEMB	1.054.389	1,29
7	COD_CENSO_INEP_MEMB	1.013.152	1,23
7	IND_CENSO_INEP_MEMB	1.004.209	1,22
7	COD_CURSO_FREQUENTA_MEMB	957.304	1,17
7	COD_ANO_SERIE_FREQUENTA_MEMB	1.430.114	1,74
7	COD_CURSO_FREQUENTOU_PESSOA_M	22.103	0,03
7	COD_ANO_SERIE_FREQUENTOU_MEMB	1.163.127	1,42
7	COD_CONCLUIU_FREQUENTOU_MEMB	435.006	0,53

### 3.1.4.9. Bloco 8 - Trabalho e remuneração

181. O Bloco 8 é de preenchimento obrigatório para as pessoas com mais de dez anos de idade e registra informações sobre a principal ocupação, quantidade de meses trabalhados no último ano, a renda auferida e o recebimento de doações, aposentadoria, seguro desemprego, pensão alimentícia ou outras fontes de remuneração.

182. A porcentagem de inconsistências nos campos varia entre 3,95% e 6,94% com média de 6%, conforme se observa na Tabela 15. Contudo, a redução na idade mínima para preenchimento do bloco, que anteriormente era de quatorze anos, explica parte das inconsistências.

Tabela 15 – Quantidade de inconsistências para o Bloco 8

Bloco	Campo	Quantidade de inconsistências	%
8	COD_TRABALHOU_MEMB	5.595.794	6,82
8	COD_AFASTADO_TRAB_MEMB	5.595.836	6,82
8	COD_AGRICULTURA_TRAB_MEMB	5.595.861	6,82
8	COD_PRINCIPAL_TRAB_MEMB	5.595.887	6,82
8	VAL_REMUNER_EMPREGO_MEMB	4.767.207	5,81
8	IND_VAL_REMUNER_EMPREGO_MEMB	4.767.207	5,81
8	COD_TRABALHO_12_MESES_MEMB	5.695.196	6,94
8	QTD_MESES_12_MESES_MEMB	5.696.463	6,94

Bloco	Campo	Quantidade de inconsistências	%
8	VAL_RENDA_BRUTA_12_MESES_MEMB	5.713.461	6,96
8	VAL_RENDA_DOACAO_MEMB	5.695.613	6,94
8	IND_VAL_RENDA_DOACAO_MEMB	5.695.613	6,94
8	VAL_RENDA_APOSENT_MEMB	3.239.899	3,95
8	IND_VAL_RENDA_APOSENT_MEMB	3.239.899	3,95
8	VAL_RENDA_SEGURO_DESEMP_MEMB	5.676.902	6,92
8	IND_VAL_RENDA_SEGURO_DESEMP_M	5.676.902	6,92
8	VAL_RENDA_PENSAO_ALIMEN_MEMB	3.869.159	4,72
8	IND_VAL_RENDA_PENSAO_ALIMEN_M	3.869.159	4,72
8	VAL_OUTRAS_RENDAS_MEMB	4.478.101	5,46
8	IND_VAL_OUTRAS_RENDAS_MEMB	4.478.101	5,46

### 3.1.4.10. Bloco 10 – Identificação de trabalho infantil

183. O Bloco informa se existe situação de trabalho infantil na família e identifica quais crianças estão nessa situação.

184. A Tabela 16 indica que a porcentagem de inconsistências nos registros é baixa, sendo inferior a 3,51% dos registros.

Tabela 16 – Quantidade de inconsistências para o Bloco 10

Bloco	Campo	Quantidade de inconsistências	%
10	IND_TRABALHO_INFANTIL_FAM	959.805	3,51
10	IND_TRABALHO_INFANTIL_PESSOA	1.858	< 0,01

### 3.1.5. Verificação da Consistência

185. A equipe de auditoria realizou testes de consistência para detectar a presença de documentos preenchidos em duplicidade.

#### 3.1.5.1. Campo CPF ou título de eleitor preenchido com valor duplicado

186. Os números de CPF e de título de eleitor são pessoais e não devem ser preenchidos em duplicidade. A Tabela 17 apresenta a situação encontrada, na qual foram detectadas duplicidade em 0,21% e 0,43% dos registros para o CPF e o Título de eleitor, respectivamente.

187. O percentual de duplicidade identificado é baixo em relação ao total de pessoas com o documento cadastrado e, em diversos casos, foram identificados a utilização do CPF do RF para o cadastramento dos demais membros da família.

Tabela 17 – Pessoas com o CPF ou o título de eleitor duplicado

Tipo de documento	Total de pessoas com o documento duplicado	Total de pessoas com o documento preenchido	%
CPF	98.453	46.892.630	0,21%
título de eleitor	146.007	34.073.980	0,43%

### ACHADO 09 – PESSOAS COM NÚMERO DE CPF E TÍTULO DE ELEITOR EM DUPLICIDADE

188. Foram identificados 98.453 registros de pessoas com número de CPF duplicado e 146.007 registros de pessoas com o título de eleitor duplicado.

#### 3.1.5.2. Documento de identidade e de carteira de trabalho preenchidos em duplicidade

189. O preenchimento do documento de identidade e da carteira de trabalho é realizado com o preenchimento de diversos campos. A comparação dos campos foi realizada a partir da concatenação dos valores preenchidos, que foram posteriormente comparados.

190. A Tabela 18 indica o total de pessoas com o documento de identidade ou carteira de trabalho em duplicidade. O percentual de pessoas com o documento de identificação duplicado é de 0,32% e para a carteira de trabalho é de 0,56%.

Tabela 18 – Pessoas com o documento de identidade ou carteira de trabalho em duplicidade

Tipo de documento	Total de pessoas com o documento duplicado	Total de pessoas com o documento preenchido	%
Doc. Identidade	147.611	46.627.613	0,32%
Cart. de Trab.	168.640	30.180.437	0,56%

#### ACHADO 10 – PESSOAS COM DOCUMENTO DE IDENTIDADE E CARTEIRA DE TRABALHO REGISTRADOS EM DUPLICIDADE

191. Foram identificados 147.611 registros de pessoas documento de identidade duplicado e 168.640 registros de pessoas com a carteira de trabalho duplicada. Para a carteira de trabalho, detectou-se diversos registros apenas com o preenchimento do estado de emissão, sem o preenchimento do respectivo número.

##### 3.1.6. Verificação da uniformidade

192. Na análise dos campos com informações quantitativas ou monetárias não foram detectados problemas de uniformidade.

##### 3.1.7. Verificação da acurácia

193. A análise da acurácia do banco de dados do CadÚnico será tratada na sessão 3.3 *Record Linkage*.

##### 3.1.8. Conclusão

194. Os testes de credibilidade desenvolvidos e aplicados na base de dados do CadÚnico permitiram uma avaliação da qualidade dos dados e do grau de credibilidade das informações produzidas no processo.

195. Os Achados 01, 02 e 03 identificaram campos do banco de dados com ausência de preenchimento significativa. Achados 04 e 05 identificaram campos que possuem as informações preenchidas, mas as regras de preenchimento entre os campos apresentam um percentual significativo de inconsistências.

196. O Achado 06 identificou os registros de indivíduos que não cumprem as regras de preenchimento do formulário para o tipo de documento que deve ser apresentado.

197. O Achado 07 identificou pessoas registradas com número de CPF e título de eleitor cujos os números informados não seguem as regras de formação. No Achado 08, foi constatado que o campo que registra o número do título de eleitor é composto por treze dígitos, enquanto o Tribunal Superior Eleitoral define o número do título com doze dígitos. Essa diferença dificulta a validação do número pela regra de formação, bem como, contribui para o registro de informações com erro.

198. Os Achados 09 e 10 identificaram pessoas com os documentos de CPF, título de eleitor, Carteira de Identidade ou carteira de trabalho com registros em duplicidade.

#### 3.2. INDICADORES GERENCIAIS

199. A elaboração de indicadores gerenciais tem o objetivo de fornecer insumos para monitorar a qualidade da base de dados e permitir a detecção de situações que ensejem a realização de fiscalizações.

200. Os indicadores foram desenvolvidos com base nos atributos de credibilidade, bem como, na escolha de variáveis importantes para a tomada de decisões. A Tabela 19 apresenta de forma resumida os indicadores elaborados.

Tabela 19 – Indicadores Gerencias

Código	Formato	Abrangência	Filtro	Descrição
01 01	Quantidade e Porcentagem	BR, Estado e Município	-	Quantidade de famílias cadastradas

Código	Formato	Abrangência	Filtro	Descrição
01 02	Quantidade e Porcentagem	BR, Estado e Município	Cadastros atualizados até 2 anos	
01 03	Quantidade e Porcentagem	BR, Estado e Município	Cadastros atualizados a mais de 2 anos	
02 01	Quantidade e Porcentagem	BR, Estado e Município	-	Quantidade de cadastros de famílias e pessoas que não seguem, pelo menos uma, das regras de preenchimento
02 02	Quantidade e Porcentagem	BR, Estado e Município	Cadastros atualizados até 2 anos	
02 03	Quantidade e Porcentagem	BR, Estado e Município	Cadastros atualizados a mais de 2 anos	
03 01	Quantidade e Porcentagem	BR, Estado e Município	-	Quantidade de cadastros de famílias e pessoas, por bloco, que não seguem, pelo menos uma, das regras de preenchimento
03 02	Quantidade e Porcentagem	BR, Estado e Município	Cadastros atualizados até 2 anos	
03 03	Quantidade e Porcentagem	BR, Estado e Município	Cadastros atualizados a mais de 2 anos	
04 01	Quantidade e Porcentagem	BR, Estado e Município	-	Quantidade de variáveis, por bloco, que não seguem a regra de preenchimento
04 02	Quantidade e Porcentagem	BR, Estado e Município	Cadastros atualizados até 2 anos	
04 03	Quantidade e Porcentagem	BR, Estado e Município	Cadastros atualizados a mais de 2 anos	
05 01	Quantidade e Porcentagem	BR, Estado e Município		Quantidade de cadastros de famílias realizados com visita domiciliar
06 01	Quantidade	BR	-	Quantidade de pessoas que o NIS repete
06 02	Quantidade	BR	-	Quantidade de pessoas que o CPF repete
06 03	Quantidade	BR	-	Quantidade de pessoas que o CPF não segue a regra de formação.
06 04	Quantidade	BR	-	Quantidade de pessoas que o título de eleitor repete
06 05	Quantidade	BR	-	Quantidade de pessoas que o título de eleitor não segue a regra de formação.
07 01	Quantidade e Porcentagem	BR, Estado e Município	-	Quantidade de pessoas que NÃO segue a regra de preenchimento de documentos (RF e demais)

### 3.2.1. Indicadores do Grupo 01

201. Os indicadores do grupo 01 informam a quantidade de famílias pelo tempo de atualização dos registros, por uf, por código de município. A partir dessas informações foram elaboradas as tabelas 19 e 20.

202. A Tabela 20 apresenta o número de famílias divididas pelo tempo de atualização do cadastro. O percentual famílias atualizadas até 2 anos é de 74% enquanto o percentual de famílias com data de atualização acima de 2 anos é de 26%.

Tabela 20 – Quantidade de famílias por data de atualização

Família	Data de atualização (X ≤ 2 anos)	Data de atualização (X > 2 anos)	Total
Baixa Renda	20.050.746	7.049.611	27.100.357
Indígena	118.836	23.288	142.124
Quilombola	120.252	15.873	136.125
<b>Total</b>	<b>20.289.834</b>	<b>7.088.772</b>	<b>27.378.606</b>

203. A tabela 21 apresenta a quantidade e o percentual de famílias por UF.

Tabela 21 – Quantidade de famílias por UF

UF	Baixa Renda	Indígena	Quilombola	Total	%
RO	243.601	2.068	193	245.862	0,90

UF	Baixa Renda	Indígena	Quilombola	Total	%
AC	122.751	3.581	2	126.334	0,46
AM	542.065	34.993	277	577.335	2,11
RR	77.730	11.899	3	89.632	0,33
PA	1.353.301	5.099	12.713	1.371.113	5,01
AP	91.689	996	586	93.271	0,34
TO	281.725	2.702	1.506	285.933	1,04
MA	1.431.353	6.117	33.147	1.470.617	5,37
PI	715.543	149	4.526	720.218	2,63
CE	1.797.746	4.132	1.691	1.803.569	6,59
RN	646.368	241	1.574	648.183	2,37
PB	805.066	4.163	2.980	812.209	2,97
PE	1.834.805	11.375	7.127	1.853.307	6,77
AL	662.515	2.937	4.664	670.116	2,45
SE	440.841	233	3.638	444.712	1,62
BA	3.027.023	7.875	36.505	3.071.403	11,22
MG	2.663.577	3.610	9.743	2.676.930	9,78
ES	459.564	706	952	461.222	1,68
RJ	1.602.182	449	1.422	1.604.053	5,86
SP	3.528.343	1.638	1.373	3.531.354	12,90
PR	1.228.678	3.889	1.240	1.233.807	4,51
SC	471.079	2.447	286	473.812	1,73
RS	1.111.993	5.885	2.894	1.120.772	4,09
MS	406.197	17.245	656	424.098	1,55
MT	478.723	7.180	1.163	487.066	1,78
GO	835.163	389	5.252	840.804	3,07
DF	240.736	126	12	240.874	0,88
<b>Total</b>	<b>27.100.357</b>	<b>142.124</b>	<b>136.125</b>	<b>27.378.606</b>	<b>100,00</b>

### 3.2.2. Indicadores do Grupo 02

204. Os indicadores do grupo 02 indicam a quantidade de cadastros de famílias e de pessoas que não seguem, pelo menos uma, das regras de preenchimento. O objetivo do indicador é identificar a proporção de registros com alguma inconsistência por município.

205. Na elaboração do indicador foram desconsiderados os campos com alto índice de inconsistências, conforme indicado nos Achados 03 e 04, pois, caso fossem considerados, iriam distorcer os resultados.

206. As Figuras 3, 4 e 5 apresentam o valor dos indicadores para os registros das famílias e as figuras 6, 7 e 8 apresentam os indicadores para os registros das pessoas. O resultado do indicador indica uma redução nas inconsistências para os registros atualizados a menos de dois anos, indicando um esforço da Unidade Jurisdicionada para reduzir a quantidade de inconsistências.

207. As Tabelas 22 e 23 apresentam os vinte municípios com os maiores percentuais de inconsistência nos registros de famílias e de pessoas, respectivamente.

Tabela 22 – Municípios com os maiores percentuais de inconsistência nos registros de família

Estado	Município	Registros com inconsistência	Total de registros	Percentual
RR	Uiramutã	1452	1468	98,9%
AM	São Gabriel da Cachoeira	7494	7577	98,9%
AM	Santa Isabel do Rio Negro	3127	3172	98,6%
RR	Normandia	1541	1612	95,6%
PB	Santa Helena	1744	1860	93,8%
MG	Lavras	11823	13217	89,5%
RR	Bonfim	2609	2927	89,1%
PB	Marcação	1764	1994	88,5%
RR	Pacaraima	2170	2476	87,6%
PB	Baía da Traição	1799	2067	87,0%
MG	São João das Missões	2695	3207	84,0%
GO	Corumbá de Goiás	2569	3204	80,2%
SC	Ipaçu	956	1222	78,2%

Estado	Município	Registros com inconsistência	Total de registros	Percentual
PR	Grandes Rios	950	1255	75,7%
GO	Itaguaru	986	1307	75,4%
PA	Jacareacanga	2071	2792	74,2%
MG	Campo Florido	1015	1393	72,9%
MG	Caparaó	689	957	72,0%
ES	São Domingos do Norte	1207	1699	71,0%
AM	São Paulo de Olivença	3425	4839	70,8%

Tabela 23 – Municípios com os maiores percentuais de inconsistência nos registros de pessoas

Estado	Município	Registros com inconsistência	Total de registros	Percentual
AP	Pracuúba	1623	2558	63,4%
RR	Normandia	4321	7536	57,3%
RS	Casca	657	1325	49,6%
AM	Manicoré	14340	29876	48,0%
RS	Boa Vista do Sul	59	126	46,8%
SC	Dona Emma	218	477	45,7%
RR	Bonfim	4147	9419	44,0%
MG	Santa Bárbara do Monte Verde	798	1946	41,0%
BA	Ipupiara	3522	8771	40,2%
SP	Pirapora Do Bom Jesus	1692	4310	39,3%
AM	Ipixuna	4301	11192	38,4%
RS	Pedro Osório	1555	4125	37,7%
MS	Itaporã	2256	6037	37,4%
SC	Capão Alto	434	1173	37,0%
RS	Sertão	700	1937	36,1%
TO	Paraíso do Tocantins	6676	18637	35,8%
SP	Itaju	479	1364	35,1%
MS	Angélica	2102	5986	35,1%
PA	Novo Progresso	4087	11803	34,6%
SP	Poloni	441	1282	34,4%

Figura 3 – Indicador 02.01. Quantidade de cadastros de famílias que não seguem, pelo menos uma, das regras de preenchimento. Todos os registros.

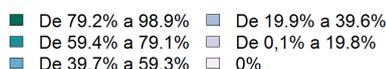
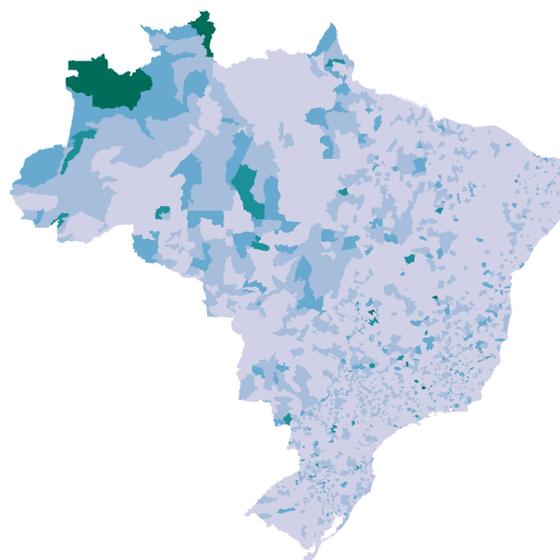
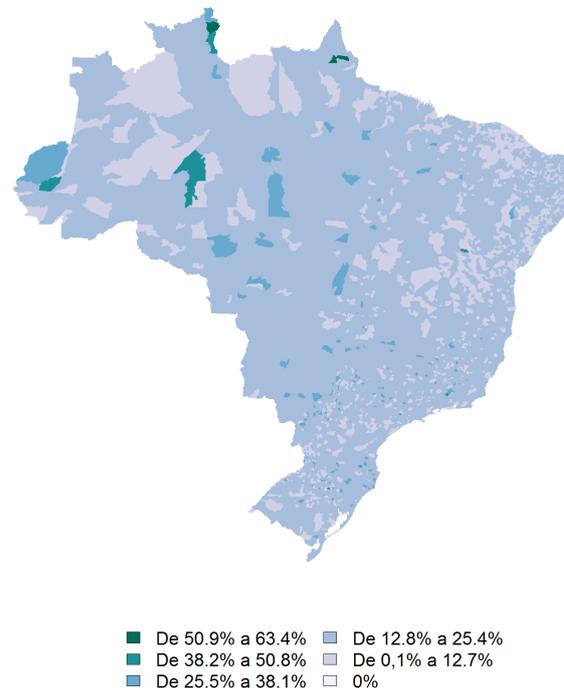


Figura 06 – Indicador 02.01. Quantidade de cadastros de pessoas que não seguem, pelo menos uma, das regras de preenchimento. Todos os registros.



### 3.2.3. Indicadores dos Grupos 03, 04 e 05

208. Os mapas com os resultados dos indicadores dos Grupos 03, 05 estão no Anexo do Relatório.

209. Os indicadores do Grupo 04 correspondem a análise dos atributos de completude e validade dos campos do CadÚnico por município. Na seção 3.1.4 foram apresentados os resultados agregados para o país, contudo, devido ao grande número de municípios não é possível inserir os resultados no Relatório. O indicadores estão no banco de dados da unidade SecexPrevi no servidor do Labcontas.

### 3.2.4. Indicadores do Grupo 06 - Documentos duplicados ou inválidos

210. Os indicadores do grupo 06 tem o objetivo de identificar se os campos destinados ao preenchimento dos documentos estão com informações duplicadas e se os números de CPF e de título de eleitor são validos. A validade dos documentos é verificada por meio da regra de formação do número e do dígito verificador. Deve-se ressaltar que, se um número cumprir a regra de formação, indica que o número pode existir, mas não significa que seja um CPF ou título de eleitor em utilização.

211. A Tabela 24 apresenta a quantidade de documentos duplicados e inválidos, cumpre ressaltar que a porcentagem de documentos com problema é baixa, com valores menores que 0,06%.

Tabela 24 – Documento duplicado ou inválido

Tipo de documento	Total de documento identificados	Total de documentos preenchido	%
CPF duplicado	98.453	46.892.630	0,21%
CPF inválido	50	46.892.630	< 0,01%
Título de eleitor duplicado	146.007	34.073.980	0,43%
Título de eleitor inválido	4.056	34.073.980	0,01%
Doc. identidade duplicado	147.611	46.627.613	0,32%
Cart. de trab. duplicada	168.640	30.180.437	0,56%

3.2.5. Indicador 07.01 - Pessoas que não seguem a regra de preenchimento dos documentos

212. O preenchimento dos campos de documentos no formulário do cadastro único possui regras específicas caso o indivíduo seja o Representante Familiar ou pertença aos demais membros da família, bem como, existem regras para famílias indígenas, quilombolas e em situação da rua.

213. O indicador 07.01 foi elaborado com base nas regras do Manual do Entrevistador e do Dicionário de Dados v5.41 (peça 09). A Tabela 25 apresenta os resultados para o estado da Bahia. Os campos de tipo de documento informam se a pessoa possui o tipo de documento preenchido. Para o caso de responsável familiar de família de baixa renda foram identificados 23.647 registros com o preenchimento indevido, nos quais, se observa o preenchimento de outros documentos, com exceção do CPF e do título de eleitor.

214. No caso dos demais membros de famílias de baixa renda foram identificados 320.943 registros com preenchimento indevido, que ocorreram, em sua maioria, devido a problemas no preenchimento dos campos de nome do cartório, número do livro, folha e termo e data do registro para as certidões de nascimento ou de casamento.

Tabela 25 – Quantidade de pessoas que não cumprem as regras de preenchimento dos documentos - Bahia

Família em situação de rua	Família	Pessoa	Quantidade	Tipo de Documento				
				Certidão	CPF	Cart. Ident	Cart. Trab	Tit. Eleit.
Não	Baixa Renda	Resp. Familiar	23.647	21.552	0	8.176	520	0
	Índigena		18	0	0	0	0	0
	Quilombola		10	0	0	0	0	0
	Baixa Renda	Demais membros	320.943	0	0	0	0	0
	Índigena		1.217	0	0	0	0	0
	Quilombola		3.722	0	0	0	0	0
Sim	Baixa Renda	Demais membros	148	0	0	0	0	0
Total			349.705	21.552	0	8.176	520	0

### 3.3. RECORD LINKAGE [comparação de registros]

215. O termo *Record Linkage* pode ser traduzido como 'comparação de registros'. Tem o objetivo de detectar entidades iguais em duas ou mais fontes de dados. Por exemplo, identificar o mesmo indivíduo em duas ou mais fontes de dados. Essa detecção é útil, pois permite averiguar se os dados de determinada base são coerentes com os dados de outras bases consideradas confiáveis. A técnica é utilizada quando não existe uma chave única de relacionamento das entidades, sendo necessário escolher os campos do registro que, quando comparados, indiquem o quanto é provável que um par de registros se refira à mesma entidade.

#### 3.3.1. Objetivo

216. Aplicar a técnica de *Record Linkage* na base de dados do CadÚnico com o objetivo de melhorar a qualidade dos dados e detectar observações com inconsistências, que podem decorrer de erros ou fraudes, ao comparar os registros dos indivíduos com os registros da base de CPF da Receita Federal.

217. As análises realizadas foram:

- a) Comparação das Bases do CadÚnico e do CPF;
- b) Enriquecimento da Base do CadÚnico.

218. Na Comparação das Bases, os registros do número de CPF e do título de eleitor foram considerados como uma chave única e os demais campos foram comparados para identificar a probabilidade das informações do CadÚnico estarem em conformidade com as informações da base de CPF da Receita Federal.

219. No Enriquecimento da base do CadÚnico, os registros dos indivíduos maiores de dezesseis anos que não possuíam o número de CPF foram comparados com as informações da base de CPF, para identificar indivíduos do CadÚnico que possuem CPF, mas não declararam durante o preenchimento do formulário de cadastramento.

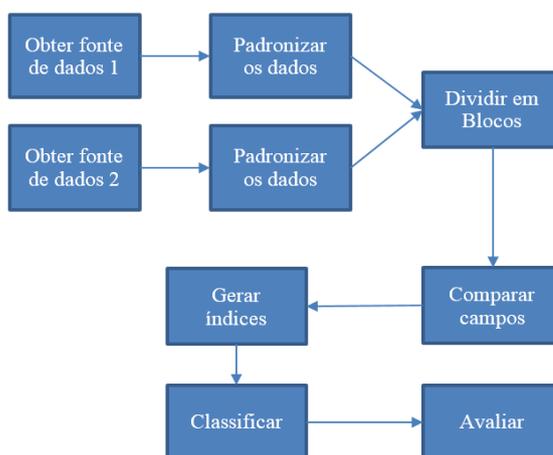
### 3.3.2. Análise de comparação das bases

#### 3.3.2.1. Metodologia

220. As principais atividades realizadas na análise de Record Linkage são apresentadas na Figura 9. A primeira atividade é a obtenção da base de dados, descrita nos itens 12 a 18. A partir da disponibilização das bases, ocorre a padronização dos campos. Os campos selecionados para a comparação das entidades, bem como, as atividades de padronização realizadas, foram:

- a) Nome. Conferir se os campos são do tipo string;
- b) Nome da mãe. Conferir se os campos são do tipo string;
- c) Data de nascimento. As datas foram convertidas para o formato 'DDMMAAAA'.
- d) Sexo. Os campos foram convertidos para valores binários, 1 para o sexo masculino e 0 para o feminino.

Figura 9 – Diagrama de blocos para a Comparação das Bases



221. A atividade de Dividir em Blocos, do termo em inglês *blocking*, consiste na divisão da base de dados em blocos, a partir da definição de um campo ou conjunto de campos, denominado chave de blocagem (*blocking key*). A chave de blocagem utilizada foi o campo CPF.

#### Comparar campos e gerar os índices

222. A comparação dos campos foi realizada pela geração de dois índices. O primeiro índice, denominado de Indicador do Power Center (IPC), obtido pela utilização das funções e algoritmos do Power Center, *software* licenciado e utilizado pelo TCU, que são capazes de comparar os registros e fornecer um resultado entre 0 e 100 para a probabilidade dos registros se referirem a um mesmo indivíduo.

223. Os manuais e instruções de uso do Power Center não descrevem em detalhes a metodologia de comparação utilizada pelo *software*. Entretanto, a leitura dos manuais e os testes realizados com o *software* permitem concluir que a comparação dos campos é realizada da seguinte forma:

- a) Campos de nome e nome da mãe. Os campos são transformados em fonemas e a distância entre os registros é computado pelo algoritmo de Jaro-Winkler;
- b) Campos de data de nascimento. A distância entre os campos utiliza o algoritmo de Humming invertido;

c) O valor do IPC é calculado com base na comparação dos campos nome, nome da mãe, data de nascimento e sexo.

224. Para validar os resultados obtidos pelo IPC, foi gerado um segundo índice, denominado Indicador de Comparação das Bases (ICB), que foi desenvolvido pela equipe de auditoria com base no modelo apresentado em Soares, 2009, e a classificação final dos registros será obtida com base nos dois índices.

#### **Indicador de Comparação das Bases**

225. O valor do ICB é obtido pela realização da comparação dos campos e pela atribuição de pesos. Os campos selecionados para a comparação e as técnicas utilizadas foram:

a) Campos de nome e nome da mãe. Os campos foram transformados em fonemas para corrigir eventuais problemas decorrentes de erros de digitação. O algoritmo utilizado foi desenvolvido a partir do algoritmo de transformação fonética desenvolvido pelo Incor/USP (<http://www.incor.usp.br/spdweb/ccsis/fonetica/>). A distância entre os campos foi calculada por meio da distância de Jaro-Winkler, que possui valores entre 0 e 100. No qual, quanto maior o valor, menor a distância entre os nomes.

b) Campo de data de nascimento. Comparadas pela distância de Humming Invertida.

c) Campo de sexo. Foi atribuído o valor 1, caso os campos sejam iguais, ou 0, caso contrário.

226. Os pesos são calculados pela comparação dos campos que compõem a entidade. Caso os campos sejam iguais ou possuam uma distância aceitável, o par de campos é considerado igual e um fator de ponderação positivo é atribuído. Caso contrário, o par é considerado diferente e um fator de ponderação negativo é atribuído. O valor final do índice corresponde a somatória dos pesos.

227. Quando um par de campos é comparado podem ocorrer quatro possibilidades:

a) Verdadeiro positivo: os campos são marcados como 'iguais' e se trata de um par verdadeiro;

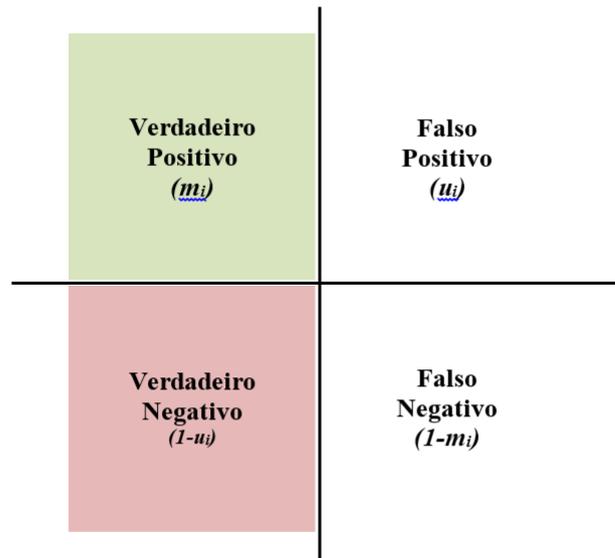
b) Falso positivo: os campos são marcados como 'iguais' e se trata de um par falso;

c) Falso negativo: os campos são marcados como 'diferentes' e se trata de um par verdadeiro;

d) Verdadeiro negativo: os campos são marcados como 'diferentes' e se trata de um par falso.

228. A Figura 10 ilustra as possibilidades que podem ocorrer ao comparar um par de campos.

Figura 10 – Possibilidades de comparação dos pares de campos



229. Considerando cada par de campos como  $i$ , o valor de  $m_i$  é definido como a probabilidade de o par ser verdadeiro positivo, a probabilidade complementar é  $(1 - m_i)$ , que corresponde a probabilidade do par ser falso negativo, ou seja, o par ser marcado como 'iguais' quando essa afirmação é falsa. O valor de  $m_i$  é chamado de sensibilidade.

230. Seja  $u_i$  a probabilidade de o par ser falso positivo, a probabilidade complementar é  $(1 - u_i)$ , que corresponde a probabilidade de o par ser verdadeiro negativo, ou seja, o par ser marcado como 'diferente' quando essa afirmação é verdadeira. O valor de  $(1 - u_i)$  é chamado de especificidade.

231. A Tabela 26 resume as informações sobre as variáveis definidas.

Tabela 26 – Quadro resumo das variáveis

Definição	Fórmula	Conceito
Sensibilidade	$m_i$	Probabilidade de o par ser verdadeiro positivo.
Especificidade	$1 - u_i$	Probabilidade de o par ser verdadeiro negativo
$(1 - \text{Sensibilidade})$	$1 - m_i$	Probabilidade de o par ser falso negativo.
$(1 - \text{Especificidade})$	$u_i$	Probabilidade de o par ser falso positivo.

232. Os valores para  $m_i$  e  $u_i$  podem ser estimados conforme as metodologias propostas por Fellegi & Sunter (1969) e Jaro (1989). Entretanto, devido ao tempo para a conclusão do trabalho, optou-se por considerar os valores utilizados por Camargo e Coli (2000) e Soares (2009), baseado nos valores estimados por Dean (1996). A Tabela 27 apresenta os valores adotados para  $m_i$  e  $u_i$ .

Tabela 27 – Valores para os parâmetros de sensibilidade e especificidade

Tipo de Campo	Sensibilidade ( $m_i$ )	1 – Especificidade ( $u_i$ )
Nome	90%	5%
Data	90%	10%
Sexo	95%	50%

233. Os valores são baseados no poder discriminatório dos campos e na probabilidade de terem seu conteúdo registrado corretamente. Para o campo nome, existe um alto poder discriminatório e considera-se que seu registro, em geral, é feito corretamente. Para o campo sexo é atribuído um baixo poder discriminatório e considera-se que seu registro esteja mais sujeito a erros.

234. Os fatores de ponderação são construídos com base nas probabilidades apresentadas. Se um par de campos for marcado como 'iguais', aplica-se o peso referente ao fator de ponderação de concordância (1):

— (1)

235. Caso contrário, aplica-se o peso referente ao fator de ponderação de discordância (2)

— (2)

236. O valor do ICB é representado pela expressão (3), que representa o valor obtido pela soma dos fatores de ponderação atribuídos a cada par de campos.

(3)

237. Na expressão (3),  $n$  é o número de campos que são comparados em cada registro,  $x$  é o fator de ponderação de concordância ou discordância e  $w$  é o peso atribuído para cada fator de ponderação.

238. Na metodologia adotada, o valor de  $m_i$  é sempre maior que  $u_i$ , dessa maneira o fator de ponderação de concordância gera um número positivo, que aumenta o valor do índice. De forma análoga, o fator de ponderação de discordância gera um número negativo e diminui o valor do índice.

239. A decisão sobre quando um par de campos deve ser considerado como 'igual' ou 'diferente' impacta diretamente no fator de ponderação aplicado. Por exemplo, o fato de o nome da mãe não ser idêntico nos campos comparados, não indica, necessariamente, que eles sejam diferentes, pois podem ocorrer uma abreviação ou a mudança do nome após o casamento. Nesses casos, é prudente adotar uma proporção mínima de concordância.

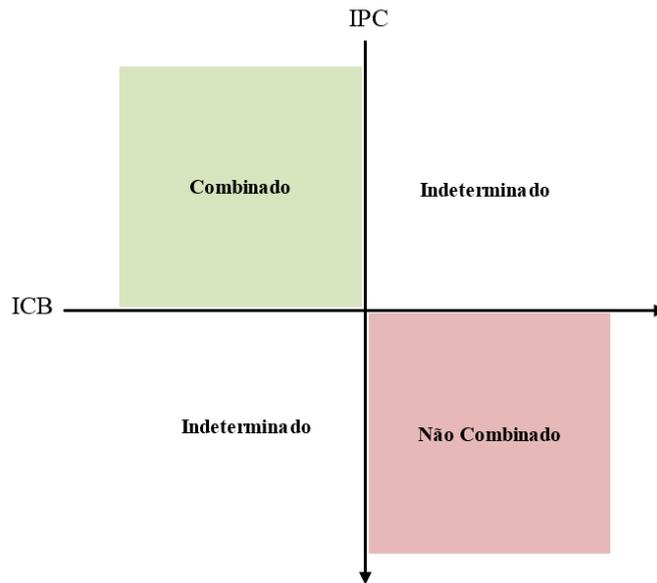
240. Com base nos testes realizados, optou-se por uma posição conservadora, que considerou como 'iguais' os campos de nome com distância de Jaro-Winkler maior ou igual a 85. Para os campos de data de nascimento, foram considerados os campos com distância de Humming Invertida maior ou igual a 89, nesses casos, aceita-se que pode ocorrer uma troca de dígito no campo de data. Por exemplo: 03/10/2010 e 30/10/2010.

### **Classificação**

241. A atividade de classificação dos registros se inicia pela escolha dos limiares superiores e inferiores, que serão utilizados para definir as categorias. Os limiares foram definidos com critérios conservadores com bases nos testes e nas análises dos algoritmos e resultados obtidos pela equipe de auditoria durante a realização das atividades de comparar e gerar os índices.

242. Os registros com valores acima do limiar superior em ambos os índices foram classificados como 'combinados' e possuem alta probabilidade de corresponderem a um mesmo indivíduo, os registros com valores menores que os limiares inferiores para os dois índices foram 'não-combinados' e possuem alta probabilidade de não corresponderem a um mesmo indivíduo. O restante dos registros foi classificado como 'indeterminado', pois os resultados obtidos não permitem afirmar com alta probabilidade se os campos correspondem, ou não, a um mesmo registro. A Figura 11 ilustra a forma de classificação.

Figura 11 – Modelo de classificação das entidades



3.3.2.2. Dados Utilizados

243. O trabalho utilizou a base do Cadastro Único referente ao mês de maio de 2015, que foi enviada pela Senarc (...) e a base do CPF de agosto de 2015 obtida por meio de acordo de cooperação entre o Tribunal de Contas da União e a Receita Federal do Brasil.

244. Conforme instruções do MDS (peça 08), somente foram considerados os registros das pessoas com estado cadastral igual a 'cadastrado' (COD\_EST\_CADASTRAL\_MEMB = 3), que resultou num universo composto por 82 milhões de registros. Não foram aplicados filtros na base de CPF que possui 214,2 milhões de registros.

**Filtro dos Registros**

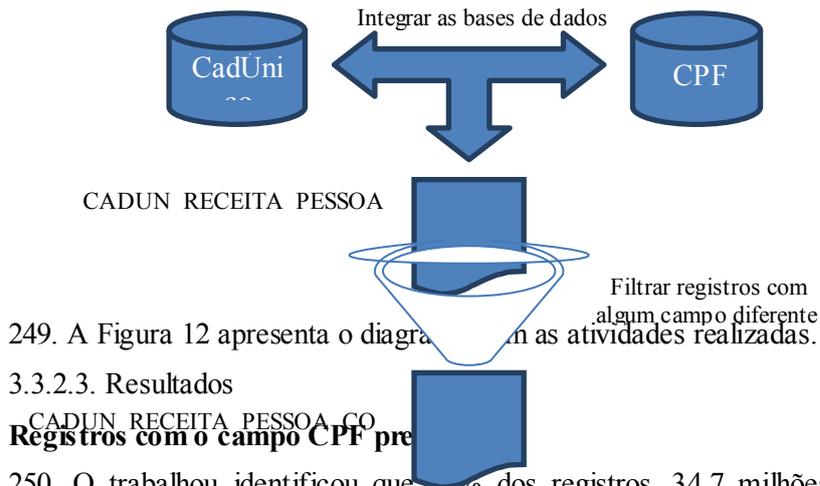
245. Os registros do CadÚnico com o campo CPF preenchido foram integrados com os registros da base do CPF, utilizando o CPF como chave única para gerar uma tabela chamada CADUN\_RECEITA\_PESSOA com os campos nome, nome da mãe, data de nascimento, sexo, CPF e título de eleitor, que possui 46.877.640 registros.

246. A comparação direta entre os pares de campos foi realizada para identificar os registros que possuíam as mesmas informações. O total de registros com todos os campos iguais é 34.706.852 (74%) e o total de registros com a algum par de campos com informações diferentes é 12.170.788 (26%).

247. Os registros com informações divergentes foram usados para gerar a base chamada CADUN\_RECEITA\_PESSOA\_COMP\_CPF, que será usada para a comparação das bases pelo CPF.

248. Os registros do CadÚnico que não possuem o campo CPF preenchido, mas possuem o campo título de eleitor preenchido foram integrados com os registros da base CPF, utilizando o título de eleitor como chave única para gerar uma tabela com os campos nome, nome da mãe, data de nascimento, sexo, CPF e título de eleitor, chamada CADUN\_RECEITA\_PESSOA\_COMP\_TE, com 905.385 registros, que será usada para a comparação das bases pelo título de eleitor.

Figura 12 – Integrar as bases e filtrar os registros



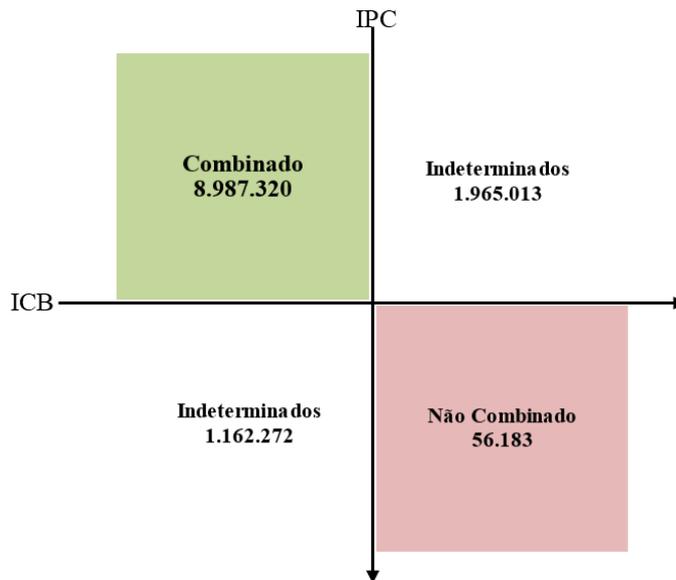
249. A Figura 12 apresenta o diagrama das atividades realizadas.

3.3.2.3. Resultados

250. O trabalho identificou que 74% dos registros, 34,7 milhões, possuem todos os campos do CadÚnico com informações iguais à base do CPF e 26% dos registros, 12,1 milhões, possuem, pelo menos, um campo com informações diferentes.

251. Os registros com informação divergente foram selecionados e comparados pelo método de *Record Linkage*. O resultado classificou com alta probabilidade 73,8% dos registros como 'combinados', 25,7% como indiferentes e 0,1% como 'não combinados', conforme apresenta a Figura 13.

Figura 13 – Resultado da comparação das bases para CPF

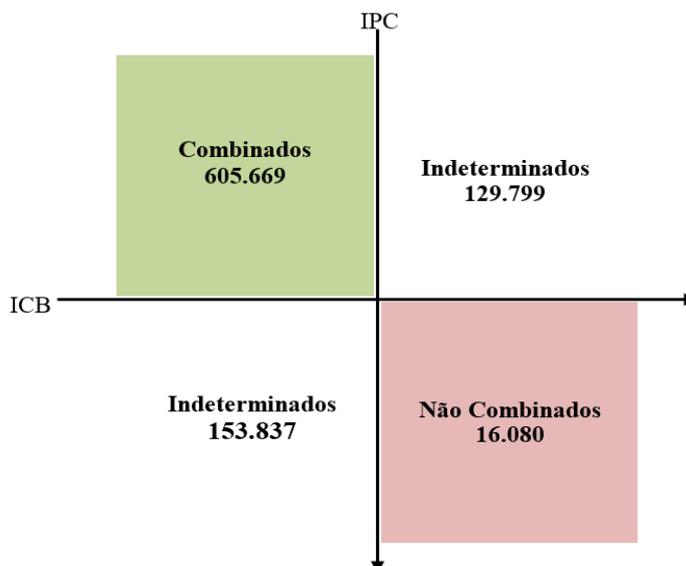


**Registros com o campo CPF nulo e o título de eleitor preenchido**

252. A base CPF possui o campo de título de eleitor preenchido para alguns registros. A fonte de origem da informação do título de eleitor é o Tribunal Superior Eleitoral, entretanto por não ter acesso a base do título de eleitor, a base do CPF foi considerada um canal confiável para a realização da comparação entre as bases do CadÚnico e do CPF para os campos de nome, nome da mãe, data de nascimento e sexo nos registros que o campo CPF era nulo e o campo título de eleitor estava preenchido.

253. Foram identificados 903.385 registros que atendiam aos critérios, dos quais 67,0% foram classificados como 'combinados', 31,4% como indiferentes e 1,8% como 'não combinados', conforme apresenta a Figura 14.

Figura 14 - Resultado da comparação das bases para o título de eleitor



### 3.3.2.4. Conclusão

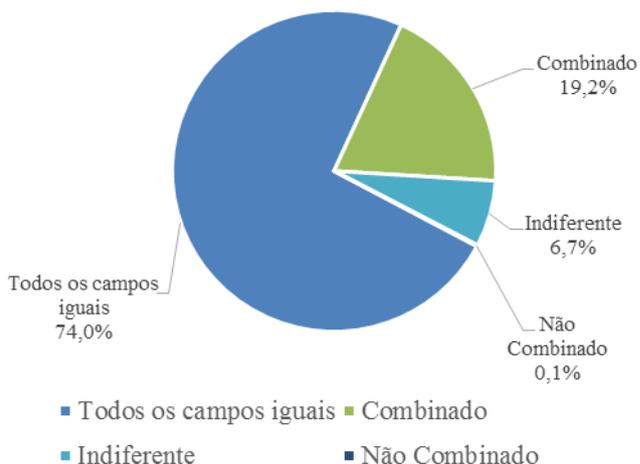
254. O objetivo da análise era comparar os pares de campos em cada registro para determinar qual a probabilidade de um mesmo registro corresponder a um mesmo indivíduo. A base de dados inicial continha 46,8 milhões de registros. A análise identificou os registros com todos os pares de campos iguais, que correspondia a 74% dos registros.

255. Os registros que não atendiam ao filtro, 26%, foram utilizados na análise de *Record Linkage* classificou os registros em 'combinados', 67,0%; indiferentes, 31,4%; e 'não combinados', 1,8%.

256. Considerando a base de dados inicial com 46,8 milhões de registros, as técnicas aplicadas foram capazes de determinar com alta probabilidade que 93,2% dos registros correspondem à um mesmo indivíduo. Desses, 74% correspondem a registros com os campos do CadÚnico com iguais aos campos da base do CPF e 19,2% correspondem a registros que foram classificados com alta probabilidade como 'combinado'. Os resultados também indicam que 0,1% dos registros não correspondem com alta probabilidade a um mesmo indivíduo, conforme indica a Figura 15

257. Os registros classificados como indiferentes, 6,7% do total, indicam que existem possibilidades de melhoria no algoritmo de classificação, bem como, possibilidades de melhoria no processo de coleta de informações e cadastramento das famílias e indivíduos.

Figura 15 – Resultado da análise de comparação das bases



### 3.3.3. Análise de enriquecimento da base do CadÚnico

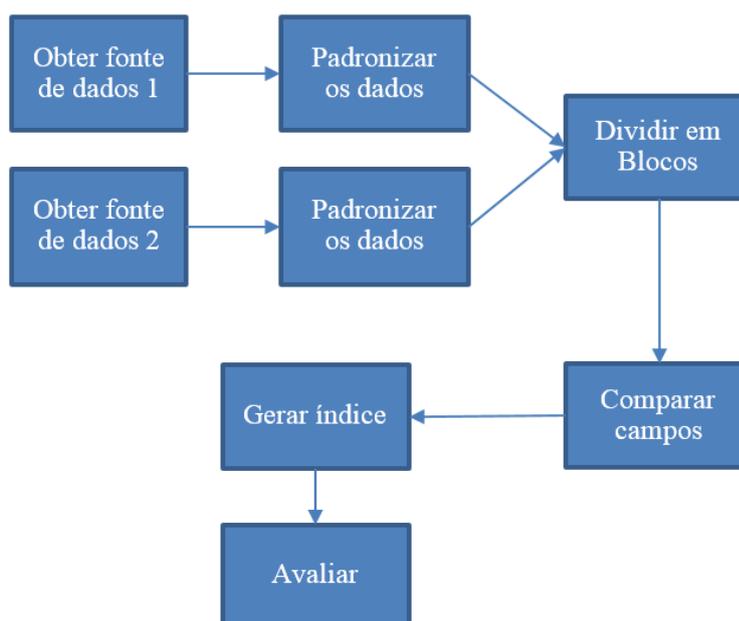
#### 3.3.3.1. Metodologia

258. A metodologia utilizada na análise de enriquecimento da base é semelhante à adotada na análise de comparação das bases.

259. As principais atividades realizadas estão apresentadas na Figura 16. A primeira atividade é a obtenção da base de dados, descrita nos itens 12 a 18. A partir da disponibilização das bases, ocorre a padronização dos campos. Os campos selecionados e as atividades de padronização realizadas foram:

- a) Nome. Conferir se os campos são do tipo *string*;
- b) Nome da mãe. Conferir se os campos são do tipo *string*;
- c) Data de nascimento. As datas foram convertidas para o formato 'DDMMAAAA'.
- d) Sexo. Os campos foram convertidos para valores binários, 1 para o sexo masculino e 0 para o feminino.

Figura 16 – Diagrama de blocos para o Enriquecimento da Base do CadÚnico



260. A Divisão em Blocos é uma das principais atividades no processo. Os registros selecionados das bases de dados não possuem uma chave única de relacionamento e caso fosse realizada uma comparação direta, seriam necessárias  $n^m$  comparações para cada par de campos, sendo  $n$  e  $m$  o número de registros em cada base.

261. A chave de blocagem utilizada foi gerada a partir da transformação da informação do campo nome em fonemas, que foram agrupados em blocos, cujo o conjunto inicial de caracteres eram iguais. Os algoritmos do *software* Power Center foram utilizados para a transformação dos nomes em fonemas e a geração dos blocos.

### **Comparar campos e gerar os índices**

262. A comparação dos campos e a geração do índice de similaridade dos registros foi realizada com a utilização dos algoritmos e funções do *software* Power Center. A equipe de auditoria optou por não desenvolver um índice próprio devido à complexidade de implementação e o prazo de realização dos trabalhos.

263. O índice de similaridade foi gerado a partir das informações dos campos de nome, nome da mãe, data de nascimento e sexo e pode assumir os valores entre 0 e 100. As comparações com valores menores que 70 foram desprezadas dos resultados.

#### **3.3.3.2. Dados Utilizados**

264. O trabalho utilizou a base do Cadastro Único referente ao mês de maio de 2015, que foi enviada pela Senarc, (...) e a base do CPF de agosto de 2015 obtida por meio de acordo de cooperação entre o Tribunal de Contas da União e a Receita Federal.

265. Conforme instruções do MDS (peça 08), somente foram considerados os registros das pessoas com estado cadastral igual a 'cadastrado' (COD\_EST\_CADASTRAL\_MEMB = 3), com idade maior ou igual a 16 e com o campo CPF não preenchido, que resultou num total de 7.939.852 registros. Não foram aplicados filtros na base de CPF que possui 214,2 milhões de registros.

#### **3.3.3.3. Resultados**

266. O trabalho para enriquecer a base do CadÚnico com o número de CPF obtido pela comparação de registros foi capaz de identificar 4,83 milhões de registros com índice de similaridade maior ou igual a 95, cujo o campo de nome da mãe esteja preenchido nas duas bases e os campos de data de nascimento sejam iguais. A quantidade de número de CPF identificados corresponde à 60,8% dos registros selecionados para a análise. A Tabela 28 apresenta a quantidade de registros para cada valor do índice de similaridade entre 100 e 95.

Tabela 28 – Resultado do Enriquecimento da Base do CadÚnico

Índice	Quantidade
100	3.997.834
99	299.643
98	118.075
97	195.527
96	171.596
95	47.615

#### **3.3.3.4. Conclusão**

267. O Enriquecimento da base do CadÚnico foi capaz de identificar com alta probabilidade 4,83 milhões de indivíduos que não declararam o número de CPF no momento do preenchimento do formulário de cadastramento. Esse número corresponde a 60,8% da quantidade inicial de registros e os resultados dessa análise serão incorporados no cruzamento de informações realizado com as bases da RAIS e do INSS para identificação de famílias e indivíduos com renda per capita superior aos limites dos programas sociais.

## **3.4. TIPOLOGIA**

### **3.4.1. Identificação de Famílias com renda per capita superior ao limite do PBF**

### 3.4.1.1. Objetivo

268. Buscar informação sobre o recebimento de rendas formais por indivíduos cadastrados no CadÚnico e beneficiários do Programa Bolsa Família. Calcular o valor da renda familiar mensal per capita e verificar quais famílias apresentam indício de fraude no PBF.

269. De acordo com o Decreto 5.209/2004, alterado pelo Decreto 8.232/2014, são famílias elegíveis para participar do PBF:

- a) Famílias com renda mensal per capita até R\$ 77,00;
- b) Famílias com renda mensal per capita entre R\$ 77,01 e R\$ 154,00 e que possuam pelo menos um membro entre 0 e 17 anos.

### 3.4.1.2. Análise do Cruzamento de Dados

#### Dados Utilizados

270. Foram utilizados os dados do CadÚnico de maio de 2015 enriquecidos com os CPF da base da Receita Federal, da folha de pagamento do Bolsa Família de junho de 2015, da RAIS de 2014 e da Maciça [banco de dados de pagamento do INSS] de julho de 2015.

#### Filtro dos Registros

271. Conforme instruções do MDS (peça 08), somente foram considerados os registros das pessoas e famílias com estado cadastral igual a 'cadastrado' (COD\_EST\_CADASTRAL\_MEMB = 3 e COD\_EST\_CADASTRAL\_FAM = 3).

272. Na folha de pagamento do PBF foram mantidos somente os registros onde pelo menos um dos benefícios recebidos está com a situação 'Liberado'.

273. Na RAIS, foram utilizados os rendimentos de 2014 dos indivíduos que mantiveram o vínculo empregatício. Também foram excluídos registros onde o rendimento de qualquer um dos meses ou do 13º salário foi maior que R\$ 72.400,00 (100 salários mínimos de 2014), assim excluindo alguns casos de lançamento de valores errados.

274. Na folha de pagamentos do INSS, foram mantidos somente os benefícios vitalícios, em geral aposentadorias por idade, invalidez e tempo de serviço e seus complementos.

### 3.4.1.3. Metodologia

#### Comparação dos Registros

275. O cruzamento dos dados do CadÚnico e das bases de renda da RAIS e da folha de pagamentos do INSS considerou o número do CPF e do PIS separadamente. Como existem casos de erros e duplicidades nestes campos, após o cruzamento dos dados foi feita uma verificação do nome e da data de nascimento dos indivíduos.

276. A comparação dos nomes utilizou sua transformação fonética e a distância de Levenshtein, sendo atribuída uma nota de 0 (completamente diferente) a 100 (exatamente iguais).

277. As datas foram primeiro padronizadas para o formato DDMMYYYY e foi utilizada a distância de Humming Invertida para dar uma nota de 0 (nenhum algarismo coincidente) a 100 (exatamente igual).

278. De acordo com a coincidência dos documentos e avaliação dos nomes e datas, os cruzamentos foram classificados em oito grupos, conforme descrito no Quadro 1.

Tabela 29 - Critérios de classificação dos indivíduos

Critério						Grupo
CPF igual	E	PIS igual	E	Nome igual	E Data igual	1
CPF igual	E	PIS igual	E	Nome > 75	E Data > 87	2
CPF igual	E	PIS igual	E	Nome > 60	E Data > 74	3
(CPF igual	OU	PIS igual)	E	Nome igual	E Data igual	4

Critério						Grupo	
(CPF igual	OU	PIS igual)	E	Nome > 75	E	Data > 87	5
CPF igual	E	PIS igual					6
(CPF igual	OU	PIS igual)	E	Nome > 60	E	Data > 74	7
(CPF igual	OU	PIS igual)					8

279. A renda da RAIS foi calculada a partir da soma dos recebimentos de janeiro a dezembro e do 13º salário dividida por 12. Dessa maneira, obteve-se a média mensal de rendimentos de 2014.

280. A renda da Maciça considerou o valor da mensalidade reajustada, recebida pelo titular do benefício em julho de 2015.

281. O número de membros da família foi calculado a partir da quantidade de NIS distintos em cada família, conforme sugerido pela documentação do CadÚnico.

282. As famílias cuja renda familiar per capita foi maior que R\$ 154,00, mas menor que R\$ 462,00, não foram consideradas como indício. Estas famílias podem ter tido sua renda estimada acima de 154 reais devido a pequenas flutuações na renda, além de sua situação de vulnerabilidade à pobreza justificar a baixa prioridade na correção dos seus benefícios.

#### 3.4.1.4. Resultados

283. Conforme se considera como certo mais grupos de cruzamento, obtém-se um número maior de famílias com indício de fraude no PBF. Na Tabela 1, são apresentadas a quantidade de famílias por grupo e o valor pago em benefícios a essas famílias.

Tabela 30 - Número de famílias com renda familiar per capita acima de R\$ 462,00 por classe de certeza sobre esta renda

Classe de certeza da família	Quantidade de famílias	Valor pago em benefícios
1	114.777	11.698.019,00
2	8.127	863.284,00
3	8.100	733.048,00
4	28.693	2.595.580,00
5	3.476	355.106,00
6	4.357	372.338,00
7	8.812	720.082,00
8	7.812	727.474,00
Total	184.154	18.064.931,00

284. Considerando as regras de criação dos grupos, espera-se que existam poucos falsos positivos nos grupos de 1 a 5, totalizando 163.173 famílias que são consideradas indícios de erro ou fraude.

285. As 20.981 famílias dos grupos de 6 a 8 serão desconsideradas nesta análise devido a maior chance de falsos positivos no cruzamento dos dados.

286. Considerando somente os cruzamentos dos grupos de 1 a 5, as famílias foram classificadas de acordo com o número de vezes que elas ultrapassam o critério do PBF, conforme a Tabela 31. Recorda-se que as famílias com renda entre 154 e 462 reais não foram consideradas como indícios.

Tabela 31 - Quantidade de famílias com indício de fraude por faixa de renda familiar per capita considerando rendas com classe de certeza 4 ou melhor

Faixa de renda familiar per capita	Número de vezes acima do critério	Quantidade de famílias	Valor pago em benefícios	Valor médio pago em benefícios
154,01 a 308,00 reais	1	629.919	91.513.730	145,28
308,01 a 462,00 reais	2	215.568	26.145.534	121,29
462,01 a 616,00 reais	3	66.887	7.568.752	113,16
616,01 a 770,00 reais	4	22.941	2.473.179	107,81
770,01 a 924,00 reais	5	51.365	4.249.043	82,72
924,01 a 1.078,00 reais	6	7.756	715.701	92,28
1.078,01 a 1.232,00 reais	7	4.598	412.147	89,64

Faixa de renda familiar per capita	Número de vezes acima do critério	Quantidade de famílias	Valor pago em benefícios	Valor médio pago em benefícios
1.232,01 a 1.386,00 reais	8	2.865	252.813	88,24
1.386,01 a 1.540,00 reais	9	1.817	158.175	87,05
Acima de 1.540,01 reais	10	4.944	415.227	83,99

287. Foi realizada uma análise das famílias com indício de erro ou fraude por entrevistador, usando a informação do CPF do entrevistador presente na Tabela Família\_01 do Cadastro Único.

288. Foram identificados 68.509 entrevistadores, inicialmente classificados de acordo com o número de famílias que estes entrevistaram, conforme a Tabela 32.

Tabela 32 - Número de entrevistadores e famílias entrevistadas e com indício de erro ou fraude por classe de famílias cadastradas pelo entrevistador

Classe de quantidade de famílias cadastradas	Total de entrevistadores	Total de famílias com indício de fraude	Total de famílias cadastradas	Percentual de indício de erro ou fraude
9 - Sem CPF	1	3.277	67.824	4,83%
8 - Mais de 2500	239	11.966	858.060	1,39%
7 - De 1001 a 2500	2.426	38.917	3.472.740	1,12%
6 - De 501 a 1000	5.130	36.414	3.588.831	1,01%
5 - De 301 a 500	5.010	22.367	1.948.501	1,15%
4 - De 176 a 300	5.977	17.548	1.385.583	1,27%
3 - De 101 a 175	6.389	12.773	859.580	1,49%
2 - De 51 a 100	7.698	9.533	559.627	1,70%
1 - De 1 a 50	35.640	10.378	476.240	2,18%
<b>Total</b>	<b>68.510</b>	<b>163.173</b>	<b>13.216.986</b>	<b>1,23%</b>

289. As 163.173 famílias com indício de fraude encontradas representam 1,23% das famílias que receberam o Bolsa Família em junho de 2015. Esta percentagem é aproximadamente a mesma nas várias classes, não evidenciando uma maior chance de fraude de acordo com o número de famílias entrevistadas por um entrevistador. No entanto, no caso das famílias onde não existe informação sobre o CPF do entrevistador, essa percentagem é muito maior.

290. Analisando os entrevistadores individualmente também foi possível identificar casos onde a taxa de famílias com indício de erro ou fraude é muito acima da média. Um exemplo com 10 entrevistadores é apresentado na Tabela 33.

Tabela 33 - Número de famílias com indício de erro ou fraude por CPF do entrevistador

CPF do entrevistador	Número de famílias com indício de fraude	Número de famílias que recebem PBF	Percentual de erro ou fraudes
293*****_**	447	3.785	11,8%
319*****_**	260	2.751	9,5%
056*****_**	226	3.887	5,8%
117*****_**	252	4.585	5,5%
153*****_**	264	5.134	5,1%
036*****_**	4	2.505	0,2%
026*****_**	4	2.637	0,2%
029*****_**	4	3.653	0,1%
540*****_**	2	3.609	0,1%
379*****_**	2	3.764	0,1%

291. Há entrevistadores com famílias com indício de erro ou fraude 10 vezes acima e abaixo da média observada no cadastro completo.

#### 3.4.1.5. Conclusão

292. Percebe-se a existência de benefícios do PBF pagos a famílias com renda familiar per capita alguma vezes maior que a regra, mas nota-se também que essas famílias conseguem obter um valor menor de benefícios. Ressalta-se o gasto de mais de R\$ 1 milhão por mês com famílias que recebem mais de mil reais por pessoa.

293. Considerando apenas as famílias com renda por pessoa acima de R\$ 462,00 e que estão nas classes de 1 a 5, foram encontradas 163.173 famílias com indício de irregularidade perfazendo um total de R\$ 16.245.037,00 de recursos do Programa Bolsa Família com indícios de pagamento indevido, para maio de 2015.

294. O cruzamento de dados foi capaz de apresentar indícios de fraude para uma quantidade expressiva de famílias e valores despendidos.

295. Há evidências de que o percentual de erros ou fraude não é homogêneo por entrevistador, indicando fragilidades no treinamento e na supervisão destes assim como a possibilidade de que alguns estejam em conluio com famílias que estejam fraudando o programa.

296. Aprimoramento nos registros de documentos no CadÚnico assim como melhorias nas regras de classificação dos cruzamentos pode aumentar o número de famílias consideradas como indício assim como diminuir a ocorrência de falsos positivos.

#### 4. CONCLUSÃO

(...)

302. Os testes de credibilidade permitiram avaliar a base de dados CadÚnico com o uso de critérios objetivos, o grau de credibilidade das informações produzida pelo processo.

303. Os registros com informações conflitantes ou em desacordo com as regras de preenchimento dos benefícios foram considerados como 'inconsistentes'. Quando o campo possuía uma quantidade significativa de inconsistências ou as inconsistências possuíam um risco maior para a ocorrência de erro ou fraude, eles eram classificados como 'achados', que serão informados ao gestor, para que, com base nos critérios de oportunidade e conveniência, indique as medidas e os controles que serão adotados (itens 138 a 193 da seção 3.1).

304. A etapa de construção de indicadores gerenciais desenvolveu dezenove indicadores gerenciais relacionados às tabelas analisadas no primeiro ciclo do acompanhamento (itens 199 a 214 da seção 3.2).

305. Os cruzamentos das bases de dados foram desenvolvidos com base na técnica de *Record Linkage* que comparou os registros dos indivíduos da base do CadÚnico com a base CPF. Na primeira análise realizada, os registros do cadastro foram comparados para identificar indivíduos com as informações da base do CPF em desacordo as informações da Receita Federal. Na segunda análise, buscou-se enriquecer a base do CadÚnico ao identificar e preencher o número de CPF para os indivíduos que não possuíam essa informação. Os resultados identificaram com alta probabilidade 72.263 indivíduos com informações inconsistentes e foi capaz de enriquecer o cadastro com o número de CPF para 4,8 milhões de indivíduos (itens 215 a 267 da seção 3.3).

306. A tipologia desenvolvida identificou famílias beneficiárias do Programa Bolsa Família com indícios de renda per capita superior ao limite. As informações da folha de pagamento do PBF foram cruzadas com as informações da RAIS e da folha de pagamento do INSS. Os resultados identificaram 163.173 famílias com indícios de erro ou fraude, cuja renda familiar é superior à quatro vezes do limite do programa, com benefícios potenciais de controle de R\$ 195 milhões (itens 268 a 296 da seção 3.4), para o ano 2015.

(...)

310. Dessa forma, verifica-se que o primeiro ciclo da Fiscalização Contínua na concessão, manutenção e pagamento de benefícios assistenciais trouxe diversos tipos de ganhos aos processos de Controle Externo realizados pelo Tribunal de Contas da União, ganhos que podem ser mantidos e maximizados pela realização de novos acompanhamentos anuais.

#### 5. PROPOSTA DE ENCAMINHAMENTO

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

- a) determinar à Secretaria Nacional de Renda de Cidadania (...) que, no prazo de 180 dias:
- a.1) Indique as providências ou os controles que serão tomados para reduzir o número de inconsistências identificadas:
    - a.1.1) Inconsistência no preenchimento do CEP e endereço (Achado 01, item 154);
    - a.1.2) Inconsistência no preenchimento do nome e do número de CPF do entrevistador (Achado 02, item 155);
    - a.1.3) Inconsistência no preenchimento do campo COD\_ESPECIE\_DOMIC\_FAM (Achado 03, item 159);
    - a.1.4) Inconsistência nas regras de preenchimento dos campos do Bloco 3 (Achado 04, itens 162 e 163);
    - a.1.5) Inconsistência nas regras de preenchimento dos campos do Bloco 4 (Achado 05, itens 166 e 167);
    - a.1.6) Registros que não cumprem as regras mínimas para preenchimento dos documentos (Achado 06, itens 171 a 174);
    - a.1.7) Pessoas com número de CPF e título de eleitor Inválido (Achado 07, item 177);
    - a.1.8) Campo do título de eleitor com número de dígitos superior ao definido pelo Tribunal Superior Eleitoral (Achado 08, item 178);
    - a.1.9) Pessoas com número de CPF e título de eleitor em Duplicidade (Achado 09, item 188);
    - a.1.10) Pessoas com documento de identidade e carteira de trabalho registrados em duplicidade (Achado 10, itens 189 e 190).
  - a.2) revise os 72.263 registros de indivíduos com as informações do CadÚnico em desacordo as informações da base do CPF da Receita Federal ou justifique a não revisão (itens 250 a 253 do Capítulo 3);
  - a.3) revise as 163.173 famílias identificadas com índice de renda per capita superior ao limite do Programa Bolsa Família (itens 283 a 286) ou justifique a não revisão;
- b) Determinar à Segecex que, no prazo de 90 dias, elabore plano de ação para:
- b.1) atualizar e disponibilizar as bases de dados e tabelas necessárias aos procedimentos desenvolvidos no âmbito do presente acompanhamento em ambiente próprio à análise da equipe de Fiscalização Contínua;
  - b.2) desenvolver novos procedimentos de Data Quality a serem integrados ao processo de Fiscalização Contínua e para criar novos painéis de visualização ou observatórios dos dados ou resultados das análises realizadas;
  - b.3) implementar as etapas de análise 'Padrões de regularidade e irregularidade', 'Construção de indicadores de resultado do negócio', 'Construção de indicador único de risco' e 'Construção de visões de risco', bem como a etapa de 'Automatização' e demais atividades relacionadas à Fiscalização Contínua;
  - b.4) desenvolver as etapas de análise já implementadas: 'Credibilidade dos dados e dos conhecimentos', 'Construção de indicadores gerenciais' e 'Tipologias', com a criação, respectivamente, de novos testes de credibilidade, indicadores gerenciais e tipologias;
- c) determinar à Secex Previdência que autue novo processo de Acompanhamento anual na concessão, manutenção e pagamento de benefícios assistenciais (Fiscalização Contínua), para o período de março de 2016 a fevereiro de 2017, em que:
- c.1) se verifiquem, para 2016, os testes de credibilidade (tópico 3.1);
  - c.2) se verifiquem, para 2016, os indicadores gerenciais (tópico 3.2);

c.3) se verifiquem novas tipologias e se desenvolvam análises de dados capazes de verificar com maior precisão a renda per capita familiar dos atuais beneficiários.

d) encaminhar ao Ministério de Desenvolvimento Social e Combate à Fome e à Secretaria Nacional de Renda de Cidadania cópia do CD com os benefícios identificados com irregularidade e com os resultados dos cruzamentos de bases, acórdão que vier a ser adotado pelo Tribunal, do relatório e do voto que o fundamentaram, bem como do inteiro teor do presente relatório;

e) Arquivar os autos."

É o relatório.