



INSTITUTO NACIONAL DE ESTATÍSTICA
STATISTICS PORTUGAL

CENSOS
2021

Estudo de viabilidade para os Censos 2021

Estudo de viabilidade da utilização de dados de fontes administrativas no novo modelo censitário para 2021
(Relatório QUAR*)

*Quadro de Avaliação e Responsabilização

9 de dezembro de 2014

Sumário executivo

No presente relatório, incorporam-se os resultados dos trabalhos desenvolvidos nos últimos meses, assim como uma reflexão sobre a margem de calendário existente para o prosseguimento da análise de fontes administrativas e opções sobre o modelo. Uma decisão sobre o modelo censitário para 2021 será tomada em 2015.

O teste a realizar em 2016, será preparado com base nessa decisão.

O relatório aborda os seguintes pontos:

- *No Ponto 1: Criação das condições legais, cooperação institucional, apoio da sociedade* – refere-se a moldura legal necessária para acesso aos ficheiros administrativos por parte do INE, com particular relevância para a Deliberação da CNPD nº 929/2014, de 11 de Junho. Identificam-se como fatores indispensáveis para o desenvolvimento do estudo a cooperação institucional com as entidades responsáveis pelos ficheiros, e o apoio da sociedade, através de uma comunicação eficaz e rigorosa.
- *No Ponto 2: Análise dos ficheiros administrativos*, são analisados 8 ficheiros administrativos de acordo com a respetiva metainformação, indicadores de qualidade e identificação das variáveis que podem ser utilizadas para o modelo censitário.

Tendo como referência as obrigações internacionais (regulamento comunitário), conclui-se que a informação administrativa permite substituir **totalmente 12 variáveis** recolhidas através dos censos, e permite substituir **parcialmente 6 variáveis**.

De um total de 77 variáveis observadas nos Censos 2011, (27 comunitárias e 50 de âmbito exclusivamente nacional) **não foi identificada informação em nenhum ficheiro administrativo que permita responder a 9 variáveis de âmbito comunitário e a 33 de âmbito nacional**.

- *No Ponto 3: Avaliação do contributo das fontes administrativas – Apresentação de resultados*, são apresentados os resultados tendo em conta a identificação das áreas não cobertas por informação administrativa,

assim como a informação que é possível utilizar e respetivas limitações. Pretende-se deste modo proporcionar um quadro de análise objetivo, ao qual se associa uma reflexão sobre hipóteses de cenários para os Censos 2021.

- *No Ponto 4: Criação da Base de População Residente (BPR) a partir de registos administrativos*, faz-se a descrição da metodologia de atualização da população residente, a partir da base de dados dos Censos 2011, assim como o modelo de infraestrutura (modelo de dados) para carregamento, integração e tratamento dos dados a partir de diferentes fontes. É realçada a necessidade de ligação entre as diferentes bases de dados e as dificuldades que esse exercício representa, tendo em conta a ausência de atributos numéricos comuns em todos os ficheiros. Foi ensaiada uma chave de matching não numérico, que permitiu colocar cerca de 8 milhões de Números de Identificação Civil na base de dados dos Censos 2011, cerca de **80%** do total de registos. Estes estudos deverão ser aprofundados e validados.
- *No Ponto 5: Avaliação do Programa de Trabalho 2013-2017 e conclusões*, é feita uma avaliação do grau de cumprimento do programa e apresentadas as conclusões. Os resultados indicam que os Censos 2021 deverão incluir a realização de um inquérito para recolher a informação não disponível em ficheiros administrativos. O desenho de um modelo censitário mais eficiente, passa pela combinação de um inquérito com informação administrativa e pela modernização do processo de recolha, no qual o ficheiro de alojamentos para fins censitários, baseado no FNA, assume um papel central.
- *No Ponto 6: Próximos desenvolvimentos e calendarização*, identificam-se as linhas de trabalho para o período 2015 / 2016 e o respetivo calendário.

Índice

Sumário executivo.....	3
Índice	5
Introdução	7
1. Criação de condições legais, cooperação institucional, apoio da sociedade	9
1.1 Condições legais.....	9
1.2 Cooperação Institucional.....	9
1.3 Apoio da sociedade	10
2. Análise dos ficheiros administrativos	11
2.1 Base de Identificação Civil (BDIC)	11
2.1.1 Breve descrição.....	11
2.1.2 Indicadores de qualidade	12
2.1.3 Identificação das variáveis (informação) que pode ser usada com fins censitários	14
2.1.4 Limitações da utilização da BDIC enquanto Base de População Residente	15
2.2 Segurança Social (SS).....	18
2.2.1 Breve descrição.....	18
2.2.2 Indicadores de qualidade	19
2.2.3 Tratamento do Ficheiro da Segurança Social.....	21
2.2.4 Variáveis que podem ser utilizadas para fins censitários	26
2.2.5 Principais limitações.....	27
2.3 Quadros de pessoal.....	27
2.3.1 Breve descrição.....	27
2.3.2 Indicadores de qualidade	28
2.3.3 Tratamento do Ficheiro dos Quadros de Pessoal	29
2.3.4 Variáveis que podem ser utilizadas para fins censitários	32
2.3.5 Principais limitações.....	33
2.4 Caixa Geral de Aposentações.....	33
2.4.1 Breve descrição.....	33
2.4.2 Indicadores de qualidade	34
2.4.3 Tratamento do Ficheiro da CGA	36
2.4.4 Variáveis que podem ser utilizadas para fins censitários	36
2.4.5 Principais limitações.....	37
2.5 Serviço de Estrangeiros e Fronteiras	37
2.5.1 Breve descrição.....	37
2.5.2 Indicadores de qualidade	38
2.5.3 Variáveis que podem ser utilizadas para fins censitários	41
2.5.4 Principais limitações.....	42
2.6 Educação.....	42
2.6.1 Breve descrição.....	42
2.6.2 Indicadores de qualidade	43
2.6.3 Tratamento dos ficheiros da Educação	45

2.6.4 Variáveis que podem ser utilizadas para fins censitários	46
2.6.5 Principais limitações	47
2.7 Imposto Municipal sobre Imóveis (IMI)	48
2.7.1 Breve descrição	48
2.7.2 Indicadores de qualidade	49
2.7.3 Identificação das variáveis que podem ser usadas para fins censitários	50
2.8 Sistema de Indicadores de Operações Urbanísticas (SIOU)	51
2.8.1 Breve descrição	51
2.8.2 Indicadores de qualidade	52
2.8.3 Identificação das variáveis que podem ser usadas para fins censitários	54
3. Avaliação do contributo das fontes administrativas - apresentação de resultados	55
4. Criação de uma Base de População Residente (BPR) a partir de registos administrativos	63
4.1 Ligação entre as várias bases de dados	63
4.2 Metodologia para a construção da Base de População Residente	64
5. Avaliação do Programa de Trabalho 2013-2017 e conclusões	75
6. Próximos desenvolvimentos e calendarização	79
ANEXOS	81

Introdução

Este relatório faz uma síntese sobre o grau de execução do Programa de Trabalho 2013-2017 - Estudo de viabilidade sobre a adoção de um novo modelo censitário.

A primeira etapa deste estudo (etapa de preparação) fica assim concluída destacando-se os seguintes pontos:

- A Deliberação da CNPD nº929/2014, de 11 de Junho, que permitiu a criação de condições legais para acesso aos ficheiros administrativos, foi um passo decisivo para o avanço do estudo. A interlocução com as diversas entidades tendo em vista a implementação da deliberação, encontra-se em curso, de modo a que o INE possa aceder a 10 bases de dados provenientes de 9 entidades diferentes;
- O estudo e a análise das experiências internacionais que permitiu a apropriação de conhecimento, fundamental para o desenho metodológico do projeto;
- A análise dos ficheiros administrativos já disponíveis permitiu identificar a informação recolhida através dos censos que pode ser substituída total ou parcialmente. Foi também possível identificar a informação não disponível em ficheiros administrativos e que tem que continuar a ser recolhida através de inquérito;
- A identificação e o dimensionamento das dificuldades associados à ligação entre ficheiros, sem atributos numéricos comuns. Nesta área foram feitos diversos ensaios, utilizando chaves de identificação não numéricas para a ligação dos ficheiros. Este trabalho permitiu aprendizagem e possibilitou resultados muito positivos;
- Foi definida a metodologia e o modelo de dados para criação de uma Base de População Residente, baseada em informação administrativa e de atualização anual.

Em 2015 o estudo entra numa segunda fase, de caracter mais operacional. Perspetiva-se que todos os ficheiros administrativos necessários se encontrem disponíveis, para que possa ser operacionalizada toda a metodologia e respetiva implementação do modelo de dados.

Antecipam-se como particularmente críticos os trabalhos de ligação de registos das bases de dados, assim como o afinamento da metodologia de atualização anual da população residente.

O contributo destes resultados, complementado com os resultados do teste de campo, a realizar em 2015, irá permitir definir o modelo de inquérito para os Censos 2021 e o respetivo contributo da informação administrativa.

1. Criação de condições legais, cooperação institucional, apoio da sociedade

1.1 Condições legais

A criação de condições legais para o acesso aos registos administrativos foi um passo decisivo para o avanço do projeto.

A par da Lei que regula o Sistema Estatístico Nacional, Lei nº 22/2008 de 13 de maio, foi necessário trabalhar com a Comissão Nacional de Proteção de Dados (CNPd) para a produção de uma deliberação que possibilitasse o acesso à informação administrativa necessária à realização do estudo de viabilidade.

A Deliberação da CNPD nº 929/2014, conjugada com a Lei do SEN, conferem a moldura legal necessária e indispensável, nesta fase. Em momento posterior e tendo em conta os resultados do estudo, a Lei de enquadramento dos Censos 2021, deverá incorporar também os princípios subjacentes na referida deliberação.

Importa contudo sublinhar que a deliberação da CNPD impõe restrições no acesso à informação administrativa, como por exemplo a limitação do acesso ao nome dos indivíduos (apenas ficarão acessíveis as 3 primeiras letras do 1º nome + as 3 últimas letras do último nome) e à respetiva morada (apenas ficarão acessíveis a Localidade e o Código Postal). Todos os identificadores numéricos (NIC, NIF, NISS) são encriptados pelas entidades antes da transmissão ao INE.

1.2 Cooperação Institucional

A cooperação por parte dos organismos detentores dos ficheiros administrativos é uma condição indispensável para o avanço do projeto. Foram identificadas, nesta fase, 9 entidades da administração pública, detentoras de informação relevante.

O GC 2021 já reuniu com todos os organismos. Nessas reuniões foi definido o conteúdo informacional, as condições de tratamento e transmissão da informação, de acordo com a deliberação da CNPD e o calendário.

O clima de grande cooperação e empenhamento revelado pelas entidades, faz antever algum otimismo na concretização do objetivo de que até ao final de 2014, todas as entidades enviem a informação ao INE.

Por proposta do INE está em apreciação da tutela, um projeto de Resolução do Conselho de Ministros, que visa fomentar as condições de cooperação e articulação entre as várias entidades envolvidas neste estudo.

1.3 Apoio da sociedade

A crescente utilização de informação administrativa e o acesso aos dados individuais, introduzem a necessidade de um maior esclarecimento junto da sociedade sobre as condições de privacidade, salvaguarda e utilização da informação.

É muito importante que a sociedade compreenda os benefícios do uso da informação administrativa como uma fonte de informação que pode substituir a informação recolhida através de inquéritos estatísticos. A redução dos custos, a melhoria da eficiência, a diminuição da carga estatística sobre os cidadãos, evitando a realização de inquéritos, a segurança dos dados, uma vez que se reduz o número de intervenientes no processo, são algumas das vantagens que devem ser percecionadas e aceites pelos cidadãos e pela sociedade em geral. Durante 2015, será dada mais visibilidade a este aspeto.

2. Análise dos ficheiros administrativos

2.1 Base de Identificação Civil (BDIC)

2.1.1 Breve descrição

A BDIC é a fonte administrativa de referência para a contabilização da população residente em Portugal e respetiva caracterização das variáveis demográficas básicas.

É constituída por todos os cidadãos de nacionalidade portuguesa residentes em Portugal ou no estrangeiro. Desde 2007, as crianças à nascença passaram a ser registadas na BDIC, ficando abrangido todo o universo da população portuguesa.

A sua utilização para fins estatísticos, apresenta contudo limitações que serão analisadas mais adiante.

A Base de Identificação Civil (BDIC) é da responsabilidade do Instituto dos Registos e Notariado (IRN) e é alimentada, pelo SIRIC (Sistema de Informação do Registo Civil que atribuía até 2007 o Bilhete de Identidade, de acordo com a *Lei nº 32/99*) e pelo Cartão do Cidadão (a partir de 2007, de acordo com a *Lei nº 7/2007*).

De acordo com informação de julho de 2014, a BDIC era constituída por cerca de **11 milhões de** indivíduos de nacionalidade portuguesa com residência em Portugal, dos quais cerca de 77% são possuidores de CC, 13,5% possuem BI vitalício e os restantes 10% possuem BI que deverão substituir pelo CC.

Quadro 1 - Indicadores estatísticos da BDIC em 6 de junho de 2014

Indicadores	Nº de indivíduos
Nº de indivíduos, na BDIC, de nacionalidade portuguesa e residência em Portugal	10 994 719
Nº de indivíduos, na BDIC, com CC de nacionalidade portuguesa e residência em Portugal	8 458 182

Indicadores	Nº de indivíduos
Nº de indivíduos, na BDIC, BI's vitalícios de nacionalidade portuguesa e residência em Portugal	1 483 401
Nº de indivíduos, na BDIC, BI's de nacionalidade portuguesa e residência em Portugal	1 053 136

Fonte: Instituto dos Registos e Notariado

Estima-se que em 2021 o número de indivíduos com CC ultrapasse os 90%, o que representa um aspeto positivo, tendo em conta a normalização e melhoria da qualidade que a base de dados do CC introduziu face à base de identificação civil, como se pode verificar através dos indicadores de qualidade apresentados no ponto seguinte.

2.1.2 Indicadores de qualidade

Os indicadores de qualidade apresentados baseiam-se na análise efetuada à BDIC 2010, uma vez que, devido a impedimentos de natureza legal o INE não teve acesso a mais nenhuma base.

A BDIC é uma base de dados consistente com elevado grau de qualidade. Globalmente as variáveis de *infraestrutura*, *apresentam boa qualidade e elevadas taxas de preenchimento*. A morada encontra-se partida, de acordo com as regras *INSPIRE* e os seus atributos *apresentam elevadas taxas de preenchimento*. Chama-se contudo a atenção para o facto de Portugal não ter um registo de moradas único e de existirem ainda muitas moradas que não possuem efetivamente todos os atributos relativos à morada. **O quadro abaixo coloca em evidência que 10% das moradas não incluem o tipo de via e a designação da via.** Não se trata de um erro de recolha, a informação não faz parte da morada. A análise destes resultados devem permitir interpretar os resultados de qualidade das moradas dos Censos 2011, bem como refletir sobre como lidar com esta situação. O código postal está preenchido com 7 dígitos e o Nome e Apelido estão normalizados e apresentam qualidade.

Quadro 2 - Qualidade de preenchimento das variáveis “infraestrutura” disponibilizadas no Ficheiro CCIC 2010 (Stock 31 de Dezembro 2010)

Nome	Taxa de preenchimento	Qualidade do preenchimento
NUM_BI	100%	Chave únicas – não existem duplicações.
CHECK_DIGITO	100%	Não existem <i>missings</i> .
NOME	100%	<i>Missings</i> (n=41); Boa qualidade de preenchimento. Todos os nomes estão em maiúsculas, com acentuação e sem abreviaturas.
APELIDO	99,82%	<i>Missings</i> (n=8 110); Os indivíduos com o campo Nome em branco (n=41) também têm o Apelido em branco; Boa qualidade de preenchimento. Todos os nomes estão em maiúsculas, com acentuação e sem abreviaturas.
NUMDI_COFRE	95,37%	<i>Missings</i> (n=206 712).
TIPO_VIA	88,04%	<i>Missings</i> (n=534 318).
DESIGN_VIA	90,48%	<i>Missings</i> (n=425 100).
TIPO_EDIFÍCIO	3,75%	<i>Missings</i> (n=4 298 566).). Não significa erro por ausência de preenchimento. Pode significar que não se aplica
PORTA	89,39%	<i>Missings</i> (n=473 801); Nos registos preenchidos aparecem situações “S/N” e “s/n”.
ANDAR	40,81%	<i>Missings</i> (n=2 643 509). Não significa erro por ausência de preenchimento. Pode significar que não se aplica
LADO	32,52%	<i>Missings</i> (n=3 013 650). Não significa erro por ausência de preenchimento. Pode significar que não se aplica

Nome	Taxa de preenchimento	Qualidade do preenchimento
LUGAR	26,77%	Missings (n=3 270 152).
LOCALIDADE	95,41%	Missings (n=204 984).
COD_POSTAL(4 dígitos)	95,83%	Missings (n=186 113);
NUM_AREA_FUNCIONA (código postal 3 dígitos)	95,83%	Missings (n=186 113); Código 000 (n= 140 923).
LOC_POSTAL	95,80%	Missings (n=187 394);

2.1.3 Identificação das variáveis (informação) que pode ser usada com fins censitários

Quadro 3 - Variáveis censitárias obrigatórias face ao regulamento comunitário possíveis de obter através da BDIC 2010

Designação da variável	Taxa de preenchimento na BDIC 2010	Observações
Local de residência habitual	100% Boa qualidade de preenchimento	Exige tratamento para aproximação do conceito da BDIC ao conceito dos censos
Sexo	100% Boa qualidade	Utilização direta
Data de nascimento	100% . Boa qualidade	Utilização direta
Estado civil legal	100% . Boa qualidade	Utilização direta
Naturalidade - Residência da mãe à data do nascimento (Município)	100% Boa qualidade	Utilização direta através da aplicação da tabela de MN às freguesias indicadas na BDIC.
Nacionalidade	100% Apenas Portuguesa e Brasileira (residual)	Apenas para a população portuguesa
Local de residência 1 ano antes	Exige estudos de análise longitudinal indivíduos em dois anos consecutivos da BDIC. Utilização pouco provável,	

Designação da variável	Taxa de preenchimento na BDIC 2010	Observações
		tendo em conta o tempo disponível.
Residência anterior no estrangeiro e ano de chegada ao país		Exige estudos de análise longitudinal dos indivíduos para vários anos da BDIC. Utilização pouco provável, tendo em conta o tempo disponível.

Quadro 4 - Variáveis disponibilizadas pela BDIC NÃO utilizadas para fins censitários

Unidade Estatística	Taxa de preenchimento na BDIC 2010	Observações
Família Clássica Relação de parentesco Pai e Mãe	Filiação através do NOME do PAI e da MÃE	A identificação dos indivíduos através do nome não faz parte da informação a disponibilizar na BDIC. Fica por isso impossibilitada a construção da relação de parentesco PAI e MÃE. Por outro lado a caracterização das relações de parentesco e a constituição de famílias e núcleos, de acordo com o regulamento comunitário tem um grau de detalhe elevado. A informação administrativa disponível não permite responder a esta temática na sua globalidade

2.1.4 Limitações da utilização da BDIC enquanto Base de População Residente

A BDIC não tem como finalidade ser uma base de população residente em Portugal. De acordo com a Lei nº 33/99 de 18 de maio que regula a identificação civil e a emissão do bilhete de identidade de cidadão nacional, a BDIC tem como finalidade, *“organizar e manter atualizada a informação necessária ao*

estabelecimento da identidade dos cidadãos e à emissão do correspondente bilhete de identidade”- Lei nº 33/99 de 18 de maio

Face ao objetivo estatístico de constituição de uma base de população residente, a BDIC apresenta dois tipos de limitações:

- **Âmbito da BDIC:** O universo da BDIC abrange apenas a população de nacionalidade portuguesa (os cidadãos brasileiros são em número residual), não inclui os cidadãos de nacionalidade estrangeira residentes em Portugal (cerca de 400 mil de acordo com os Censos 2011).
- **Conceito entre população registada (BDIC) e população residente (Censos)**

De acordo com o conceito censitário população residente (independentemente da nacionalidade) é aquela que vive ou tem a intenção de viver no seu local de residência habitual (local onde reside com a família e tem a maior parte dos seus haveres, independentemente do momento censitário) por um período contínuo de pelo menos 12 meses.

De acordo com a BDIC, a população registada deve indicar livremente uma morada que corresponde ao local de residência onde pode ser regularmente contactado. A atualização da morada deve ser feita logo que deixe de ser possível o seu contacto regular no local anterior.

A diferença conceptual entre a BDIC e os Censos, determina diferenças muito significativas no apuramento da população residente de acordo com a BDIC 2010 e os Censos 2011 **(cerca de 1,1milhões de pessoas a mais na BDIC 2010)**. Estas diferenças podem ser explicadas porque efetivamente há indivíduos que indicam uma morada de contacto em Portugal, mas à luz dos censos não são residentes. Por outro lado há ainda que ter em atenção os fluxos de população em território nacional e que muitas vezes não são acompanhados da respetiva atualização da morada junto dos organismos competentes, o que origina distorções na contabilização da população residente ao nível regional e com maior impacto para níveis geográficos mais desagregados.

Quadro 5 - Diferenças, por NUTS II entre o nº de pessoas (nacionalidade portuguesa) na BDIC 2010 e nos Censos 2011

NUT II	BDIC 2010 (Nº)	Censos 2011 (Nº)	BDIC - Censos (Nº)	BDIC/Censos (%)
Total	11 231 695	10 145 862	1 085 833	10,7
Norte	4 027 703	3 631 075	396 628	10,9
Centro	2 560 350	2 268 547	291 803	12,9
Lisboa	2 816 064	2 611 736	204 328	7,8
Alentejo	767 774	732 253	35 521	4,9
Algarve	432 103	397 882	60 026	8,6
R.A. Açores	302 813	242 787	60 926	24,7
R.A. Madeira	324 888	261 582	63 306	24,2

Muitos países da UE, possuem um registo da população independente da Base de dados de Identificação Civil. Estes registos são geridos e atualizados ao nível municipal, sob a coordenação dos Institutos de Estatística. A título de exemplo refere-se o caso espanhol com o *Padrón Municipal*. Os países que possuem registos de população municipal, criados precisamente para determinar a população residente no município, têm também problemas de atualização que resultam em diferenças entre a contabilização da população residente (conceito estatístico) e a população registada (conceito *dos ficheiros administrativos*). Essas diferenças são contudo mais fáceis de ultrapassar nos países com registos de população ao nível do município do que em Portugal, em que apenas se dispõe da BDIC.

A construção de uma metodologia que possibilite a transformação da BDIC num ficheiro estatístico de população residente representa um dos principais desafios deste estudo de viabilidade e uma das condições necessárias para alteração da mudança do modelo censitário, através da incorporação de informação administrativa.

Para uma análise mais desenvolvida sobre a BDIC 2010 deve ser consultado o relatório “Análise de ficheiros administrativos BDIC 2010, de 02-12-2013”.

2.2 Segurança Social (SS)

2.2.1 Breve descrição

A Base de Dados de Qualificação da Segurança Social, foi disponibilizada pelo Instituto de Informática da Segurança Social e segundo a metodologia de extração de dados definida, corresponde às pessoas que têm uma qualificação ativa (válida), no sentido de um enquadramento no sistema de segurança social.

A informação constante nesta Base de Dados da Segurança Social abrange todas as pessoas singulares que mantêm uma inscrição ativa na segurança social, independentemente do tipo de qualificação: trabalhador por conta de outrem, trabalhador independente, pensionista de velhice, pensionista de invalidez, rendimento social de inserção,....

Os dados da Segurança Social permitem uma aproximação à caracterização de uma parte muito significativa da população empregada, parte da população desempregada e parte da população que recebe reforma ou outras prestações sociais, conforme já identificado em relatórios anteriores¹.

Não fazem parte do âmbito deste ficheiro, as pessoas que não integram o Regime Geral de Segurança Social, nomeadamente:

- as pessoas que se encontram ao abrigo de outros regimes de proteção social, como por exemplo, os trabalhadores da administração pública com entrada até 2006;
- os trabalhadores que exercem atividade independente e cujo patamar de rendimento seja inferior a 6 vezes o IAS (em 2013 o IAS era 419,22€) ou cuja duração da atividade seja inferior a 12 meses;
- os reformados que recebem reformas ou outras prestações sociais, ao abrigo de outros regimes, como por exemplo da Caixa Geral de Aposentações.

¹ Avaliação dos ficheiros administrativos: Segurança Social e Quadros de Pessoal , Dezembro de 2013

Não constam igualmente nesta base de dados, as pessoas que, por algum motivo não tenham qualquer ligação ativa à SS na data de referência.

O potencial de utilização do ficheiro da Segurança Social em termos de aproveitamento para fins censitários assenta em duas linhas:

- contribui de forma decisiva para a aferição da população residente em termos censitários, na medida em que abrange uma parte muito significativa da população residente em Portugal;
- é o ficheiro central na componente relativa ao emprego e mercado de trabalho (apesar de não abranger toda a população nem todas as variáveis censitárias).

2.2.2 Indicadores de qualidade

A informação fornecida pela Segurança Social revela-se fundamental não só pelo número de indivíduos que abrange, mas também pelo facto de agregar numa única base de dados os 3 identificadores numéricos mais relevantes: NISS - Número de Identificação da Segurança Social, NIC - Número de Identificação Civil e NIF - Número de Identificação Fiscal. Esta última característica faz deste ficheiro uma plataforma crucial na articulação da informação proveniente das várias entidades.

O identificador numérico principal da Segurança Social é o NISS, o qual tem uma taxa de preenchimento de 100%. Destaca-se ainda com um bom nível de preenchimento o NIF com cerca de 98%. Em contrapartida, apenas 76% dos registos da SS possuem NIC (5 419 937), o que pode condicionar um emparelhamento mais célere e seguro com as restantes fontes administrativas.

A comparação dos NIC's entre a Segurança Social e a BDIC permitiu constatar que dos 5 419 937 NIC's existentes na SS 5 186 629 têm correspondência na BDIC, ou seja cerca de 96%, o que atesta a qualidade do preenchimento deste campo na presente base de dados.

As variáveis relacionadas com o local de residência habitual (Morada, Código postal, localidade,...) apresentam um preenchimento de 98%. Contudo, o código postal apresenta um número considerável de códigos apenas com 4 dígitos.

Para a freguesia de residência a taxa de preenchimento é de 96%, enquanto o município e o distrito de residência tem um preenchimento na ordem dos 98%.

Quadro 6 – Indicadores de qualidade das “variáveis de estrutura” da Segurança Social, ano 2013

Número de registos únicos	7 124 233	
Designação da variável	Taxa de preenchimento	Observações
NISS	100%	
NIC	76%	5 419 937 NIC's preenchidos dos quais têm correspondência na BDIC 5 186 629 (96%)
NIF	98%	
Nome	100%	
Distrito de residência	98%	
Município de residência	98%	
Freguesia de residência	96%	
Morada – artéria	98%	Morada não particionada. Embora formato seja passível de normalização.
Localidade	98%	
Código postal	98%	Alguns apenas com 4 dígitos

A análise das bases de dados da SS para os anos de 2008, 2009, 2010, 2011 e 2013 evidencia que a estrutura da BD e as taxas dos indicadores da qualidade do preenchimento se mantêm estáveis, sendo que em algumas situações estas taxas melhoram consideravelmente. Para o ano de 2013, destaca-se a melhoria na taxa

de preenchimento do NIC e do NIF. Os desenvolvimentos e conclusões dos vários trabalhos foram apresentados em relatórios próprios.

2.2.3 Tratamento do Ficheiro da Segurança Social

Não sendo um ficheiro estatístico, a informação da segurança social foi tratada no sentido de fornecer dados estatísticos, com possibilidade de serem utilizados para efeitos censitários.

A Segurança Social disponibilizou informação sobre todas as qualificações ativas que a pessoa possuía à data de referência (31 de Dezembro do ano de referência), o que originou que cada pessoa pudesse ter que mais do que uma qualificação - situação existente para quase 2 milhões de registos.

No sentido de assegurar a existência de apenas uma qualificação para cada indivíduo, foi implementado um conjunto de regras, cujo princípio geral foi o da prevalência das qualificações relacionadas com atividade económica sobre as situações de reforma, apoio social, etc, princípio equivalente ao utilizado para a informação de carácter censitário.

Para os indivíduos que tinham mais do que uma qualificação relacionada com atividade económica, por exemplo 2 ou mais situações de trabalho por conta de outrem ou trabalho por conta de outrem e trabalho de forma independente, foram tomadas um conjunto de decisões, destacando-se por exemplo a existência de correspondência nos quadros de pessoal.

Posteriormente foi construída uma tabela de correspondência, que permitiu transformar/aproximar as qualificações da segurança social, em modalidades da variável censitária condição perante a atividade económica, criando-se 4 grandes categorias: população empregada, população desempregada, população reformada e outra situação (onde se incluiu por exemplo os indivíduos que

recebem Rendimento Social de Inserção ou outro tipo de apoios não passíveis de classificação nas categorias anteriores)².

A comparação dos resultados entre a Segurança Social e Censos, para o ano de 2011, mostra que os censos têm valores mais elevados em qualquer das categorias, como seria expectável, uma vez que o universo da SS não é exaustivo.

A diferença entre os Censos 2011 e a SS2011 na população empregada é cerca de 198 mil indivíduos (4,5%), na população reformada de 260 mil (11,1%).

A divergência mais significativa refere-se à população desempregada, onde a diferença entre as duas fontes é de quase 50%. Note-se que na Segurança Social apenas é possível classificar como desempregados as pessoas que recebiam prestações relacionadas com o desemprego.

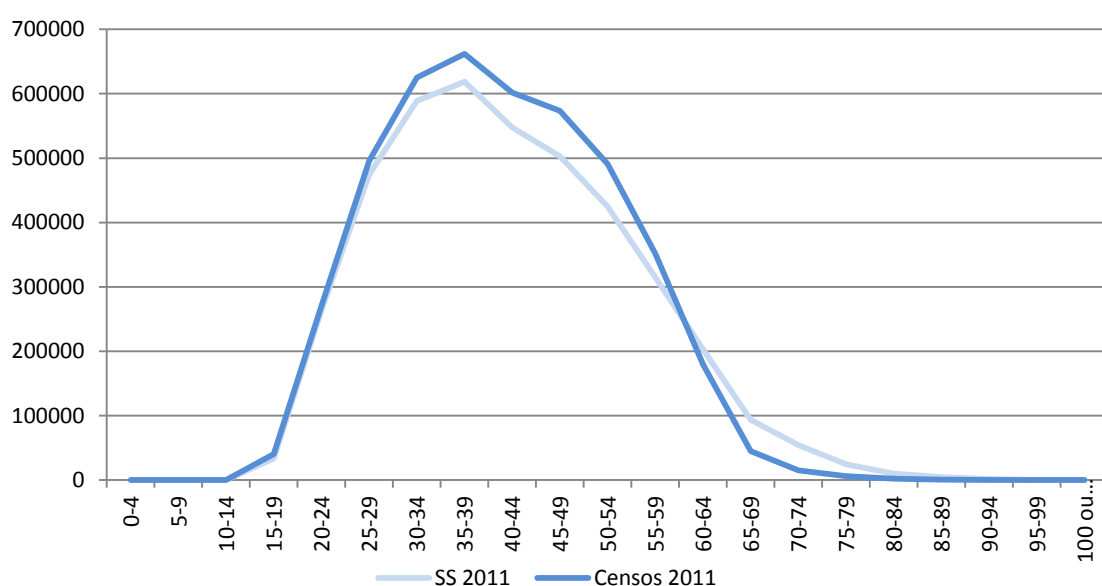
Quadro 7 – População empregada, desempregada e reformada – Segurança Social 2011 versus Censos 2011

Condição perante a atividade económica	Censos 2011	SS 2011	Diferença absoluta	Diferença relativa %
População empregada	4 361 187	4 163 045	198 142	4,5
População desempregada	662 180	333 912	328 268	49,6
População reformada	2 339 094	2 078 385	260 709	11,1

² Esta tabela de correspondência é provisória, uma vez que ainda existem algumas dúvidas sobre algumas qualificações da Segurança Social.

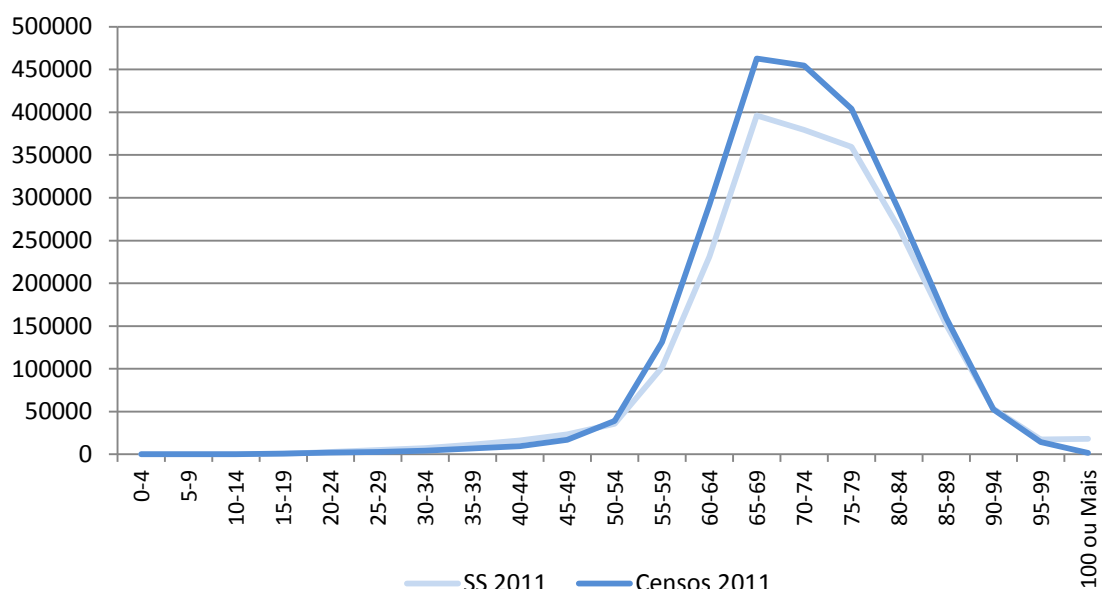
Para o ano de 2011, a comparação da população empregada por grupo etário mostra uma distribuição relativamente próxima entre as duas fontes, ainda que seja evidente um valor mais elevado de população empregada nos censos para os grupos etários entre os 35 e os 55 anos e a inversão desta tendência a partir dos 64 anos.

Fig 1 – População empregada por grupo etário, Segurança Social 2011 versus Censos 2011



Também a comparação por grupo etário da população reformada evidencia uma consistência entre as duas fontes. As diferenças mais significativas, com valores mais elevados nos Censos 2011, estão concentradas nas idades entre os 65 e os 79 anos.

Fig 2 – População reformada por grupo etário, Segurança Social 2011 versus Censos 2011



Da análise por município, ressalta que a grande maioria dos municípios tem mais população empregada nos censos do que na segurança social, sendo Coimbra o município onde esta diferença é maior em termos absolutos. Em sentido contrário, aparece Lisboa, como o município onde os Censos 2011 apuraram menos empregados que os dados da Segurança Social.

Quadro 8 – População empregada, Segurança Social 2011 versus Censos 2011, Município

População empregada	Nº de municípios	Maiores diferenças absolutas		Maiores diferenças relativas (%)	
Diferenças positivas (Mais nos Censos que na SS)	283 (92%)	Coimbra	9 443	Barrancos	26,3
		Oeiras	7 107	Tarouca	22,2
		Vila Nova de Gaia	6 020	Boticas	21,9
		Vila Franca de Xira	5 352	Vila Nova de Paiva	21,1
		Maia	5 123	Valpaços	21,0
Diferenças negativas	25	Lisboa	-5 418	Câmara de Lobos	-10,4

População empregada	Nº de municípios	Maiores diferenças absolutas		Maiores diferenças relativas (%)	
(Menos nos Censos que na SS)	(8%)	Cascais	-2 143	Póvoa de Lanhoso	-8,8
		Câmara de Lobos	-1 386	Ribeira Brava	-4,2
		Loulé	-880	Vila Viçosa	-3,89
		Póvoa de Lanhoso	-719	Vieira do Minho	-3,71

Relativamente à população reformada, também a esmagadora maioria dos municípios tem mais população reformada nos Censos do que na Segurança Social. Lisboa, Oeiras e Sintra apresentam as maiores diferenças absolutas, acima dos 10 mil indivíduos. De assinalar ainda, que em alguns municípios estas diferenças ultrapassam os 30%, conforme quadro seguinte.

Quadro 9 - População reformada, Segurança Social 2011 versus Censos 2011, Município

População Reformada	Nº de municípios	Maiores diferenças absolutas		Maiores diferenças relativas (%)	
Diferenças positivas (Mais nos Censos que na SS)	296 (96%)	Lisboa	11 264	Melgaço	35,9
		Oeiras	10 934	Almeida	34,9
		Sintra	10 783	Aljezur	34,6
		Cascais	9 665	Murtosa	32,2
		Almada	8 807	Sabugal	31,7
		Porto	8 469	Lagos	30,2
Diferenças negativas (Menos nos Censos que na SS)	12 (4%)	Santo Tirso	-521	Porto Moniz	-19,5
		Arouca	-422	Arouca	-9,1
		Feira	-206	Câmara de Lobos	-3,3
		Câmara de Lobos	-139	Santo Tirso	-3,1
		Porto Moniz	-127	São João da Madeira	-2,6

2.2.4 Variáveis que podem ser utilizadas para fins censitários

A informação fornecida pela Segurança Social revela-se fundamental pela qualidade, pela dimensão da população abrangida, e pelo facto de ter associado os 3 identificadores numéricos mais relevantes.

O quadro seguinte sintetiza as variáveis da Segurança Social que podem ser utilizada para fins censitários:

Quadro 10 – Variáveis da Segurança Social que podem ser utilizadas para fins censitários

Designação da variável	Observações
Morada, Localidade Freguesia e município de residência	Preenchimento a 98%. Contributo para a variável local de residência
Sexo	Preenchimento a 100%. Variável auxiliar
Data de nascimento	Preenchimento a 100%. Variável auxiliar
Estado civil	Preenchimento a 98%. Variável auxiliar, embora face ao estado civil legal acresça a identificação das uniões de facto.
País de naturalidade	Preenchimento a 100% Variável auxiliar
Freguesia e município de naturalidade	Preenchimento a 100% (Para os registos com país de naturalidade Portugal) Variável auxiliar
País de nacionalidade	Preenchimento a 91%. Variável Auxiliar
Qualificação	Preenchimento a 100%. Esta variável é a que define o tipo de relacionamento do Indivíduo com a SS (se é trabalhador por conta de outrem, trabalhador independente, se recebe uma pensão de reforma, se recebe uma prestação de desemprego ou de apoio social, etc..). Serve de base à construção da variável censitária “Condição perante a atividade económica”, nomeadamente População empregada,

Designação da variável	Observações
	desempregada e reformada, embora não abranja a totalidade das situações. A qualificação fornece também informação sobre a variável Situação na profissão para TPCO, TPCP e Patrões (aproximação).
NPC da empresa	Preenchimento a 100%. Pode servir de base à variável censitária Ramo de atividade económica da empresa, e à variável número de trabalhadores da empresa (não obrigatória em termos internacionais) através de ligação ao FUE.
Morada da empresa (Código de distrito município freguesia, morada)	Preenchimento a 100%. Serve de fonte à variável censitária Local de trabalho.

2.2.5 Principais limitações

- O âmbito do ficheiro, que apesar de abranger uma parte muito expressiva da população, não fornece informação para o universo censitário;
- Sendo o ficheiro da SS um elemento fundamental na articulação entre as diferentes fontes, o nível de preenchimento do NIC (na ordem dos 76%), representa um constrangimento ao condicionar a possibilidade de ligação por chave numérica de uma forma mais abrangente;
- A atualização do ficheiro da segurança social, nomeadamente a atualização das qualificações.

2.3 Quadros de pessoal

2.3.1 Breve descrição

Os Quadros de Pessoal, Anexo A do Relatório Único, são disponibilizados pelo Gabinete de Estratégia e Planeamento do Ministério da Solidariedade, Emprego e

Segurança Social e resultam da obrigação anual, a cargo dos empregadores, do envio de informação sobre a atividade social da empresa.

A informação disponibilizada pelos Quadros de Pessoal abrange as pessoas empregadas com contrato individual de trabalho, em empresas com pelo menos uma pessoa ao serviço e abrangidas pelo código do trabalho. Assim, estão fora do âmbito dos quadros de pessoal os trabalhadores por conta própria, os trabalhadores do serviço doméstico, os trabalhadores da administração pública, dos institutos públicos e os que possuem regime de contrato de trabalho em funções públicas.

Na perspetiva de aproveitamento de informação administrativa para efeitos censitários, os Quadros de Pessoal funcionam como informação complementar aos dados da Segurança Social, na área do emprego e mercado de trabalho, possibilitando uma caracterização mais alargada do grupo dos trabalhadores com contrato individual de trabalho.

2.3.2 Indicadores de qualidade

A análise das “variáveis de estrutura” mostra que o único identificador numérico associado ao indivíduo nos Quadros de Pessoal é o NISS, o qual se encontra preenchido a 100%.

O matching através do NISS, entre os quadros de pessoal e a base da Segurança Social, para o ano de 2011, foi cerca de 96%, o que indica uma aderência muito forte entre estes dois ficheiros. Ainda assim, em cerca de 110 mil registos (4%) não houve correspondência entre as duas bases, o que pode indiciar eventualmente pequenos problemas no preenchimento do NISS ou desajustamento das datas de referência (QP outubro, SS dezembro).

Quadro 11 – Indicadores de qualidade das “variáveis de estrutura” dos Quadros de Pessoal, ano 2011

Número de registos únicos	2 735 523	
Designação da variável	Taxa de preenchimento	Observações
NISS	100%	O matching com o ficheiro da SS 2011 é de 96%.
NIF da empresa	100%	
Morada da empresa (Código de distrito município freguesia, morada , localidade...)	100%	

As variáveis de caracterização do indivíduo (sexo, data de nascimento, profissão, habilitações literárias, ...) e as relativas às empresas (ramo de atividade económica, número de trabalhadores) estão preenchidas praticamente na totalidade.

Para as variáveis que utilizam “classificações nacionais” como profissão e atividade económica, verificou-se que os códigos eram válidos.

2.3.3 Tratamento do Ficheiro dos Quadros de Pessoal

Decorrente do facto de existirem pessoas que trabalham em mais do que uma empresa (2,2% do universo dos QP) e tendo em conta que para efeitos censitários apenas se caracteriza o “emprego principal”, a informação dos quadros de pessoal foi sujeita a um conjunto de regras de tratamento no sentido de assegurar apenas um registo por indivíduo.

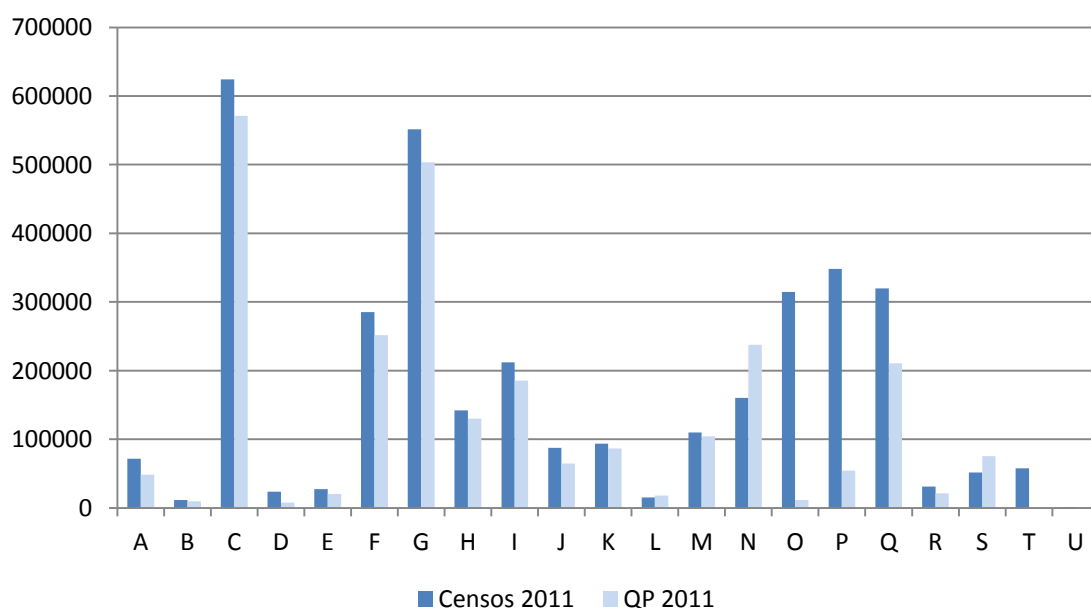
De uma forma geral, para os indivíduos com mais de uma ocorrência a determinação do registo a manter foi realizada de acordo com o maior número de

horas de trabalho na semana, critério idêntico ao utilizado habitualmente nos censos da população.

A comparação entre os Censos 2011 e os Quadros de Pessoal para a população empregada por conta de outrem (situação na profissão) dá conta de uma diferença de 928 mil indivíduos a menos (26%) nos Quadros de Pessoal. A não inclusão no universo dos Quadros de Pessoal dos “funcionários públicos”, dos trabalhadores domésticos, entre outras situações podem justificar esta diferença.

A análise dos ramos da atividade económica vem evidenciar isso mesmo; Comparando com os resultados dos Censos 2011 as secções da CAE com menor representatividade nos QP são, como seria expectável, a Administração Pública e Defesa (secção O), Educação (secção P), Saúde (secção Q) e as atividades das famílias empregadoras de pessoal doméstico (secção T)³.

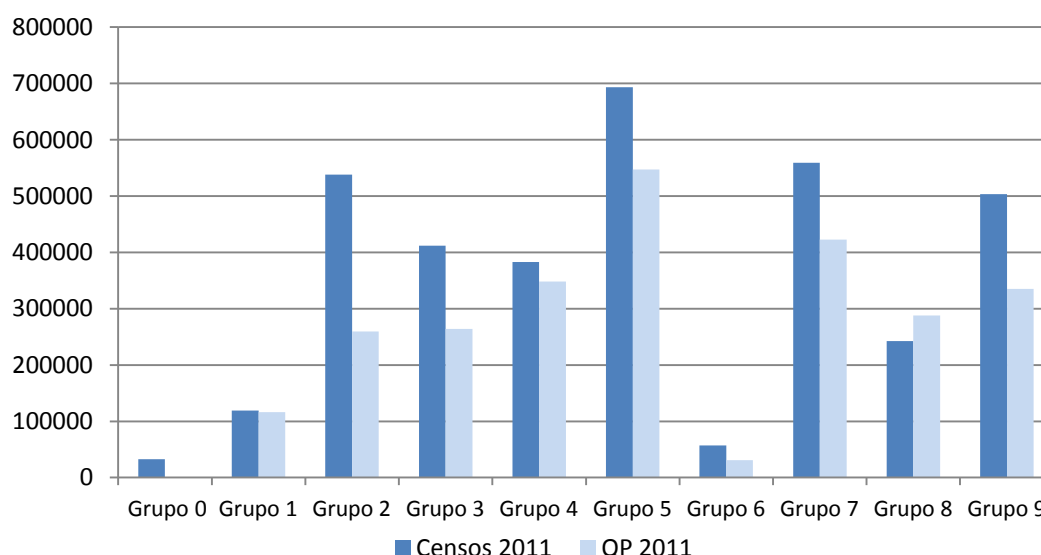
Fig 3 - Trabalhadores por conta de outrem, segundo a secção de atividade económica, QP 2011 e Censos 2011



³ Para uma análise mais detalhada ver *Análise dos Ficheiros administrativos: Emprego e Mercado de trabalho - Relatório de Progresso, Julho de 2014*

Na comparação das profissões a 1 dígito, para o universo dos trabalhadores por conta de outrem, a diferença mais significativa encontra-se no Grupo 2 - *Especialistas das atividades intelectuais e científicas*, com uma diferença na ordem dos 278 mil indivíduos. Também no Grupo 9 - *Trabalhadores não qualificados*, Grupo 5 - *Trabalhadores dos serviços pessoais, de proteção e segurança e vendedores*, Grupo 3 - *Técnicos e profissões de nível intermédio* e Grupo 7 - *Trabalhadores qualificados da indústria, construção e artífices* as diferenças são consideráveis, e variam entre os 168 mil e os 136 mil indivíduos.

Fig 4 - Trabalhadores por conta de outrem, segundo o grupo profissional, QP 2011 e Censos 2011



- Grupo 0 Profissões das Forças Armadas
- Grupo 1 Representantes do poder legislativo e de órgãos executivos, dirigentes, diretores e gestores executivos
- Grupo 2 Especialistas das atividades intelectuais e científicas
- Grupo 3 Técnicos e profissões de nível intermédio
- Grupo 4 Pessoal administrativo
- Grupo 5 Trabalhadores dos serviços pessoais, de proteção e segurança e vendedores
- Grupo 6 Agricultores e trabalhadores qualificados da agricultura, da pesca e da floresta
- Grupo 7 Trabalhadores qualificados da indústria, construção e artífices
- Grupo 8 Operadores de instalações e máquinas e trabalhadores da montagem
- Grupo 9 Trabalhadores não qualificados

2.3.4 Variáveis que podem ser utilizadas para fins censitários

O quadro seguinte identifica as variáveis relacionadas com a área do emprego e mercado de trabalho, que podem ser aproveitadas para fins censitários, tendo presente que apenas caracterizam parte da população empregada, nomeadamente a abrangida pelos Quadros de Pessoal.

Quadro 12 – Variáveis dos Quadros de Pessoal que podem ser utilizadas para fins censitários

Designação da variável	Observações
Situação na profissão	Preenchimento a 100%. Disponibiliza informação para patrões (desde que fazem parte dos quadros de pessoal da empresa), TPCO, membros ativos de cooperativa e trabalhadores familiares não remunerados. Serve como fonte para a variável censitária Situação na profissão.
Profissão (CPP 5 dígitos)	Preenchimento a 100%. Codificada a 5 dígitos da CPP2010 Serve como fonte para a variável censitária profissão.
Ramo de atividade económica da empresa	Preenchimento a 100%. Codificado a 5 dígitos da CAE rev.3 Serve como fonte para a variável censitária atividade económica da empresa/estabelecimento.
Ramo de atividade económica do estabelecimento	Preenchimento a 100%. Codificado a 5 dígitos da CAE rev.3 Serve como fonte para a variável censitária atividade económica da empresa/estabelecimento.
Morada da empresa (Código de distrito município freguesia, morada , localidade...)	Preenchimento a 100%. Serve como fonte para a variável censitária Local de trabalho.
Morada do estabelecimento (Código de distrito município freguesia, morada , localidade...)	Preenchimento a 100%. Serve como fonte para a variável censitária Local de trabalho.
Número de pessoas ao serviço da empresa	Preenchimento a 100%. Serve como fonte para a variável censitária (não obrigatória em termos de regulamento comunitário) Número de pessoas ao serviço na empresa/estabelecimento.
Número de pessoas ao serviço no estabelecimento	Preenchimento a 100%. Serve como fonte para a variável censitária (não obrigatória em termos de regulamento comunitário)

Designação da variável	Observações
	Número de pessoas ao serviço na empresa/estabelecimento.
Período normal de trabalho semanal	Preenchimento a 100%. Serve como fonte para a variável censitária (não obrigatória em termos de regulamento comunitário) Número de horas de trabalho na semana.
Habilitações literárias	Preenchimento a 99,7% (Classificação que sintetiza o nível de ensino possuído e a área de formação no caso do ensino superior). Pode afigurar-se como fonte para a variável censitária nível de ensino completo.

2.3.5 Principais limitações

- O principal constrangimento da informação disponibilizada pelos quadros de pessoal é o seu âmbito: o facto de estar restringido apenas à população empregada com contrato individual de trabalho, condiciona uma utilização mais eficaz desta informação para o universo censitário;
- Para o ano de 2011 não foi disponibilizada a informação relativa aos Açores.

2.4 Caixa Geral de Aposentações

2.4.1 Breve descrição

A Caixa Geral de Aposentações (CGA) é a instituição de previdência que tem a seu cargo a gestão do regime de segurança social dos funcionários públicos e trabalhadores equiparados admitidos até 31 de dezembro de 2005 em matéria de pensões de aposentação, de reforma, de sobrevivência e de outras de natureza especial⁴.

⁴ Até 31 de Dezembro de 2005, eram obrigatoriamente inscritos na CGA os trabalhadores da Administração Pública Central, Regional e Local e de outras entidades públicas, que tivessem a qualidade de funcionários públicos ou agentes

A informação disponibilizada pela CGA abrange os indivíduos subscritores (funcionários públicos e agentes administrativos civis e militares da Administração Pública Central, Local e Regional, professores do ensino particular e cooperativo e trabalhadores de algumas empresas públicas e sociedades anónimas de capitais públicos), os reformados da função pública e os pensionistas de sobrevivência de funcionários públicos ou equiparados.

Os dados da CGA apresentam especial interesse como fonte complementar ao ficheiro da Segurança Social, quer na caracterização das pessoas que já não se encontram no ativo, mas também dos funcionários da administração pública com data de admissão anterior a 2006.

2.4.2 Indicadores de qualidade

Na informação disponibilizada pela CGA, para além do identificador principal - Número de subscritor – constam também os identificadores numéricos NIC e NIF.

O número de subscritor encontra-se preenchido na totalidade (1 161 304 registos), enquanto o NIC tem um grau de preenchimento de 71% (826 mil registos). Por sua vez, o NIF apresenta uma taxa de preenchimento bastante mais baixa, apenas 47% (547 mil).

Para as variáveis relacionadas com o local de residência (distrito e município de residência, morada e código postal) verifica-se uma fraca taxa de preenchimento: apenas metade dos registos possuem essa informação.

administrativos. Desde 1 de Janeiro de 2006, o pessoal admitido na função pública passou a ser inscrito no regime geral da segurança social.

Quadro 13 – Indicadores de qualidade das “variáveis de estrutura” do ficheiro CGA – ano 2008

Número de registos únicos	1 161 304	
Designação da variável	Taxa de preenchimento	Observações
Nº de subscritor	100%	
BI (NIC)	71%	Dos 826 677 NIC's preenchidos fazem <i>matching</i> com a BDIC 93,5%.
NIF	47%	
Distrito de residência	50%	
Município de residência	50%	
Morada	50%	Morada não particionada;
Localidade	18%	
Código postal	50%	A maioria dos registos tem preenchimento a 7 dígitos

Existem alguns problemas na uniformização e harmonização dos campos, por exemplo, a tabela de distrito/município utilizada não é a que está atualmente em vigor.

A determinação exata do universo abrangido unicamente pela CGA tem-se revelado difícil, uma vez a sobreposição com a Segurança Social não se conseguiu aferir de forma precisa (por diversas situações a mesma pessoa pode estar referenciada nas duas bases de dados, por exemplo a pessoa é reformada da Segurança Social e pensionista de sobrevivência da CGA,...)

A comparação entre as duas bases de dados é dificultada pelas taxas de preenchimento dos identificadores numéricos comuns, pelo desajustamento das datas de referência e pelo facto do ficheiro da CGA atualmente disponível não ter nomes. Assim, e tendo como referência o NIC, cerca de 27% dos registos com BI na CGA (cerca 220 mil) estão também na SS; para o NIF cerca de 42% dos registos com NIF (cerca de 229 mil) estão na SS. Estes valores devem ser lidos com alguma cautela, devido às taxas de preenchimento destas variáveis.

2.4.3 Tratamento do Ficheiro da CGA

O ficheiro da CGA possuía duplicados ao nível do número de subscritor, consequência do mesmo indivíduo poder acumular diferentes situações: por exemplo ser pensionista de sobrevivência ou subscritor e outra situação. Neste sentido foram implementadas um conjunto de regras de modo a assegurar que, para cada número de subscritor existe apenas um registo. Na definição das regras deu-se prevalência às situações de atividade económica, de acordo com as regras habitualmente utilizadas em termos censitários.

Paralelamente existiam também números de subscritores distintos para o mesmo número de BI. Também aqui se procedeu à eliminação de duplicados, tendo cada número de BI ficado associado a um único número de subscritor.

2.4.4 Variáveis que podem ser utilizadas para fins censitários

A grande utilidade do ficheiro da CGA é a complementaridade, em termos de universo, que proporciona ao ficheiro da Segurança Social. Em termos de variáveis, não há muita informação que se possa traduzir em variáveis censitárias, no entanto, a variável Situação na CGA fornece algumas indicações sobre a situação do indivíduo face à atividade económica.

Quadro 14 - Variáveis CGA que podem ser utilizadas para fins censitários

Designação da variável	Observações
Situação na CGA	100% de preenchimento. Fornece indicações sobre a condição do indivíduo perante a atividade económica (Subscritor - empregado, reformado, outra situação)
Sexo	100% de preenchimento. Variável auxiliar
Data de nascimento	100% de preenchimento. Variável auxiliar

2.4.5 Principais limitações

As principais limitações da informação proveniente da CGA são:

- Fracas taxas de preenchimento dos identificadores numéricos BI e NIF, que comprometem a análise de complementaridade entre este ficheiro e a Segurança Social;
- Ausência de nome, o que inviabilizou o matching através de chave de ligação, quer com a Segurança Social, quer com a BDIC. Provavelmente numa próxima disponibilização esta limitação possa ser ultrapassada;
- Fraca taxa de preenchimento das variáveis relacionadas com o local de residência, apenas metade dos registos na CGA disponibiliza esta informação.

2.5 Serviço de Estrangeiros e Fronteiras

2.5.1 Breve descrição

O Serviço de Estrangeiros e Fronteiras (SEF) é o organismo responsável pelo registo da informação relativa à entrada, permanência e afastamento de cidadãos estrangeiros em território Nacional.

O ficheiro do SEF contém informação sobre os indivíduos de nacionalidade estrangeira com título de residência válido, e pode permitir o recenseamento, em termos de cobertura e de algumas variáveis de conteúdo da população estrangeira residente legalmente em Portugal.

A informação proveniente do SEF é essencial para completar a população residente em Portugal, uma vez que a informação constante da BDIC abrange apenas a população com nacionalidade portuguesa.

2.5.2 Indicadores de qualidade

O Ficheiro do SEF para o ano de 2011 contabiliza um total de 437 729 indivíduos de nacionalidade estrangeira, valor superior em cerca de 10% (40 mil) quando comparado com os resultados dos Censos 2011.

A análise da informação disponibilizada pelo SEF evidencia, de uma forma geral, bons indicadores de qualidade para o conjunto de variáveis demográficas básicas (sexo, data de nascimento, pais de nacionalidade, país de naturalidade).

Atendendo às variáveis de “estrutura”, a taxa de preenchimento do Número de Autorização de Residência é de 99,5%. No entanto, para os identificadores numéricos que permitem a interligação com outros dados administrativos, NIF e NISS, as taxas de preenchimento são bastante baixas, não atingindo em nenhum dos casos 50%. O NIF tem uma taxa de preenchimento de 47,9% (208 mil registos) e o NISS de 39,6% (173 mil registos).

Outra limitação é ainda o facto destes identificadores NIF e NISS estarem associados a mais de uma pessoa. O mesmo número pode estar associado a todo o agregado familiar.

As variáveis morada e localidade têm uma taxa de preenchimento na ordem dos 99,5%, embora no caso específico da morada, algumas não são indicadas de forma precisa e referem apenas o nome do lugar/localidade, bairro ou nome da rua.

O código postal a 4 dígitos encontra-se preenchido em 99,3% dos registos, no entanto a taxa de preenchimento decresce ligeiramente (93%) tendo como referência o código postal a 7 dígitos (4 + 3).

Quadro 15 – Indicadores de qualidade das “variáveis de estrutura” do ficheiro do SEF

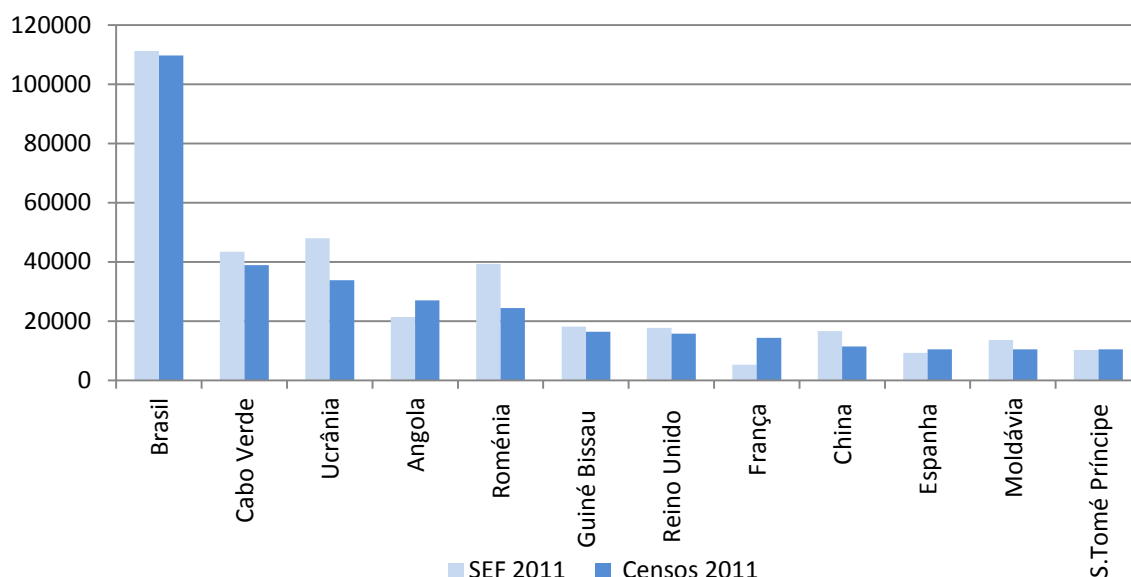
Número de registos únicos	437 729	
Designação da variável	Taxa de preenchimento	Observações
Nome	100%	
Número de AR	99,9%	
NIF	47,6%	O mesmo NIF está associado a várias pessoas do agregado familiar.
NISS	39,6%	O mesmo NISS está associado a várias pessoas do agregado familiar.
Distrito de residência	100%	
Município de residência	100%	
Morada	99,5%	Morada não particionada; Algumas moradas não são precisas Ex: apenas o nome do lugar
Código postal - 4 dígitos	99,3%	
Código postal - 3 últimos dígitos	93,0%	
Localidade	99,5%	

Os dados disponibilizados pelo SEF têm um universo superior em 10% (40 mil) face aos resultados dos Censos 2011, no entanto, essa diferença não se distribui de forma uniforme pelas diferentes nacionalidades.

A comparação da população estrangeira por nacionalidade evidencia em termos gerais uma estrutura próxima, ainda que seja possível distinguir algumas nacionalidades onde as diferenças são mais significativas.

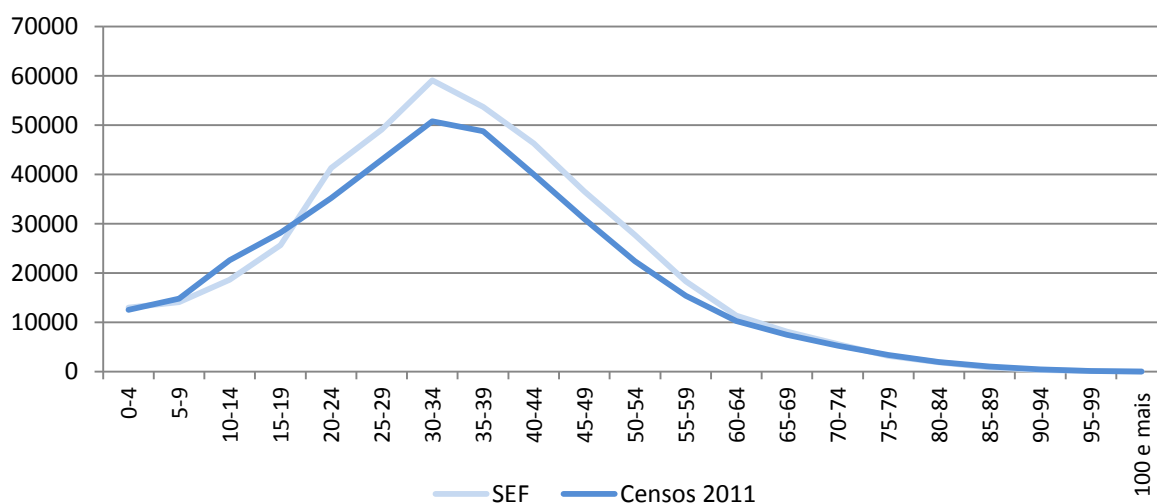
Na população de nacionalidade romena e ucraniana as divergências são na ordem dos 14 mil indivíduos a mais no SEF, enquanto para a comunidade francesa a diferença, de cerca de 9 mil indivíduos, é em sentido contrário (a menos no SEF comparativamente aos Censos 2011).

Fig 5 – População de nacionalidade estrangeira, segundo o país de nacionalidade, SEF 2011 e Censos 2011



A comparação da população estrangeira por grupo etário mostra uma sobrerrepresentação no SEF por referência aos censos para a maior parte das idades, com exceção das idades compreendidas entre os 5 e os 19 anos.

Figura 6 – População de nacionalidade estrangeira, segundo o grupo etário, SEF 2011 e Censos 2011



2.5.3 Variáveis que podem ser utilizadas para fins censitários

As variáveis que o SEF disponibiliza podem ser utilizadas para caracterização demográfica da população estrangeira residente. Apesar de conter também informação sobre a situação perante a atividade económica e a profissão, o seu grau de preenchimento é bastante baixo o que condiciona a caracterização da população estrangeira face ao mercado de trabalho.

Quadro 16 - Variáveis SEF que podem ser utilizadas para fins censitários

Designação da variável	Observações
Morada (morada, código postal, localidade, Distrito e Município de residência de residência)	99% de preenchimento. Não tem nível geográfico freguesia. Fonte para a variável Local de residência
Sexo	100% de preenchimento
Data de nascimento	100% de preenchimento
País de Nacionalidade	100% de preenchimento
País de Naturalidade	100% de preenchimento
País de última residência	28% de preenchimento
Situação profissional	37% de preenchimento Fornece indicação sobre a condição perante a atividade económica da pessoa (Ativo/Não ativo)
Tipo de ocupação	37% de preenchimento Fornece indicação sobre a condição perante a atividade económica da pessoa (Empregado/ desempregado/ reformado/ estudante/...) e a Situação perante a profissão (TPCO, TPCP,...)
Grupo profissional	37% de preenchimento. Profissão a 1 dígito CNP 94
Profissão	20% de preenchimento. Profissão a 6 dígitos CNP 94 Pode servir de fonte à variável censitária Profissão.

2.5.4 Principais limitações

- O Ficheiro do SEF abrange apenas os indivíduos estrangeiros residentes com título de residência válido;
- Fraca taxa de preenchimento dos identificadores numéricos que permitem a ligação com outras fontes administrativas, nomeadamente NISS e NIF;
- A atualização do ficheiro, nomeadamente o abate dos estrangeiros que saem do país com o título de residência ainda válido;
- Ausência do nível geográfico freguesia de residência (*esta questão poderá ser ultrapassada com o tratamento de moradas/códigos postais*);
- As variáveis de caracterização da população face ao mercado de trabalho têm taxas de preenchimento muito baixas.

2.6 Educação

2.6.1 Breve descrição

No final do 1º trimestre de 2014, o INE recebeu, pela primeira vez, da Direção Geral de Estatísticas da Educação e Ciência (DGEEC) dados individualizados relativos a alunos matriculados/inscritos no Ensino Básico, Secundário e Superior.

Segundo foi apurado, a informação disponibilizada é proveniente de várias fontes nomeadamente:

- Recenseamento escolar anual, dirigido a estabelecimentos de educação e ensino privados e públicos de outros Ministérios;
- Dados reportados pelos estabelecimentos de educação e ensino públicos;
- Dados provenientes do Sistema de Informação e Gestão da Oferta Educativa e Formativa (SIGO), Instituto de Emprego e Formação Profissional, Centros de Novas Oportunidades.

Os dados da educação vão permitir caracterizar a população que se encontra a frequentar o ensino, no sentido da sua incorporação como estudante na condição perante o mercado de trabalho.

Outro grande propósito deste ficheiro é a sua utilização na validação da população residente para fins censitários, atendendo a que produz indícios de residência muito consistentes.

Importa salientar, que os dados da educação não permitem assinalar os níveis educacionais da população residente no seu todo, mas apenas quantificar e caracterizar os indivíduos que estão a frequentar o ensino.

2.6.2 Indicadores de qualidade

Em relação aos indicadores de qualidade dos dados da educação para as “variáveis de estrutura” salienta-se o seguinte:

- Apesar do preenchimento quase total dos campos tipo de documento de identificação e número do documento de identificação, apenas 83% corresponde a Número de Identificação Civil BI/CC (cerca de 1 580 000), destes 88% fazem matching com a BDIC (1 393 000);
- Falta de normalização na utilização do NIC (registos apenas com o número de identificação, outros com o número propriamente dito + o check digit, e ainda outros com Numero, check digit e caracteres alfanuméricos); - A identificação através do número da cédula pessoal/certidão de nascimento abrange cerca de 190 000 registos, o que condiciona a comparação por matching numérico;
- O município e o distrito de residência são desconhecidos para mais de 300 mil registos;
- Fraco preenchimento dos campos relativos a freguesia de residência, códigos postais e morada: cerca de 700 mil registos não tem freguesia de residência, para 535 mil não consta o código postal e em mais de 778 mil a morada não está preenchida.

Quadro 17 – Indicadores de qualidade das “variáveis de estrutura” dos dados da educação - ano letivo 2011/2012

Número de registos	1 909 939	
Designação da variável	Taxa de preenchimento	Observações
Nome	100%	
Tipo de documento de identificação	98%	83% corresponde a NIC (cerca de 1580000); 10% corresponde a número de cédula pessoal/ certidão de nascimento. (cerca de 190000) e os restantes distribuem-se entre Autorizações de residência, Passaporte e outros.
Número do documento de identificação	98%	83% corresponde a NIC (cerca de 1580000); 10% corresponde a número de cédula pessoal/ certidão de nascimento. (cerca de 190000) e os restantes distribuem-se entre Autorizações de residência, Passaporte e outros. Falta de normalização do NIC em relação à utilização do check digit e caracteres alfanuméricos.
Distrito de residência	84%	
Município de residência	84%	
Freguesia de residência	63%	
Morada	59%	Morada não particionada;
Código postal	72%	Falta de normalização em relação à utilização do código postal a 7 dígitos

2.6.3 Tratamento dos ficheiros da Educação

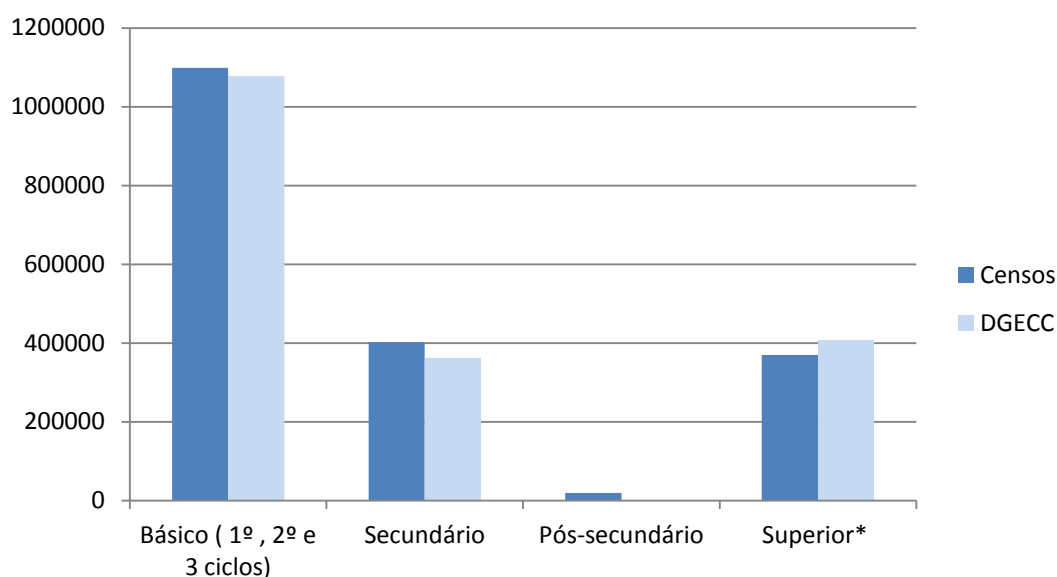
A análise dos dados relativos à educação foi bastante dificultada pela falta de harmonização da informação recebida, quer em termos de estrutura de dados quer em termos de codificação de variáveis. Os dados relativos à educação foram recebidos em 7 tabelas distintas, correspondendo a níveis e modalidades de ensino diversas.

Procedeu-se à compatibilização da informação e construiu-se uma única base de dados para o ano letivo 2011/2012, onde se agregou a informação recebida para um conjunto de variáveis consideradas básicas.

Outra dificuldade encontrada relaciona-se com a existência de registos duplicados de difícil identificação/eliminação. Exemplo: Existem situações em que a mesma pessoa tem documentos de identificação distintos - exemplo cédula e BI; Em algumas situações para o mesmo número de BI o nome não está descrito exatamente da mesma forma;.....).

Apesar de não ser possível aferir com exatidão (devido à existência de duplicados, registos onde o nível de ensino não está discriminado,...) podemos considerar os dados recebidos pela educação muito consistentes com os Censos 2011 ao nível dos grandes totalizadores. Os Censos 2011 apuraram, para o Continente, 1 890 090 indivíduos a frequentar o ensino e os dados da DGEEC 1 909 996, perfazendo uma diferença de apenas 19 906 indivíduos (cerca de - 1,1%).

Fig 7 – Alunos a frequentar o sistema de ensino, Continente (não inclui pré-escolar), DGECC versus Censos 2011



Nota: No Ensino Superior os dados da Educação incluem as regiões autónomas.

2.6.4 Variáveis que podem ser utilizadas para fins censitários

O grande potencial da informação disponibilizada pela DGECC, mais do que as variáveis propriamente ditas é o universo abrangido.

Quadro 18 – Variáveis da Educação que podem ser utilizadas para fins censitários

Designação da variável	Observações
População que frequenta o ensino	Após tratamento para Estudante (frequenta o ensino e não trabalha) é uma das categorias para a construção da variável censitária condição perante a atividade económica.
Sexo	100% de preenchimento. Variável auxiliar
Data de nascimento	100% de preenchimento. Variável auxiliar
País de Nacionalidade	100% de preenchimento. Variável auxiliar

Designação da variável	Observações
Nível de ensino frequentado	97% de preenchimento. Serve de fonte para a variável Nível de ensino a frequentar/frequentado. Pode servir de base à variável censitária Nível de ensino concluído, tendo como premissa que a frequência de um nível de ensino pressupõe a conclusão do anterior.

2.6.5 Principais limitações

A análise do ficheiro da educação evidenciou os seguintes constrangimentos:

- Para o ensino Básico e Secundário, o âmbito geográfico abrangido é apenas para Portugal Continental. Apenas para o Ensino Superior o âmbito geográfico é Portugal;
- Uma outra limitação, ligada ao âmbito do ficheiro, foi a não disponibilização de informação relativa ao ensino pré-escolar;
- Falta de harmonização da informação recebida ao nível de estrutura e no preenchimento de algumas variáveis;
- Fraco preenchimento dos campos relativos a morada, códigos postais e códigos de freguesia;
- A identificação através do número da cédula pessoal/certidão de nascimento para cerca de 10% dos registos, o que condiciona a comparação, através de matching numérico, com outras fontes administrativas, nomeadamente a BDIC.

2.7 Imposto Municipal sobre Imóveis (IMI)

2.7.1 Breve descrição

O IMI entrou em vigor em 1 de dezembro de 2003, nos termos do Decreto-Lei nº 287/2003, de 12 de novembro. Substituiu a anterior Contribuição Autárquica, em vigor desde 01.01.1989.

O IMI incide sobre o valor patrimonial tributário dos prédios rústicos e urbanos situados no território português. Para os efeitos do presente relatório, interessa considerar alguns conceitos e disposições, incluídos no Código do IMI, que se apresentam no anexo 3.

Cada andar ou parte de prédio suscetível de utilização independente é considerado separadamente na inscrição matricial. Assim, no contexto deste relatório, o termo “prédio urbano” pode equivaler a “moradia”, a “fração autónoma” ou a “alojamento familiar”.

No processo de inscrição ou atualização de prédios urbanos, é registado um conjunto de elementos de caracterização, do titular e do prédio, parte importante dos quais é suscetível de utilização na produção estatística. Mais à frente são apresentados os elementos que interessa considerar no âmbito das variáveis censitárias.

Na sequência de protocolo assinado em 2005, o INE recebe da Autoridade Tributária (AT), informação sobre o IMI, em ritmo anual. Através de uma 2ª adenda a esse protocolo (uma 1ª adenda, em 2007, refere-se ao IMT), foi estabelecido em 2012 o envio mensal da informação relativa ao IMI, necessária à atividade do INE, mantendo-se em paralelo os envios anuais.

Decorre atualmente o processo de revisão global do protocolo de 2005, tendo sido proposta a inclusão das variáveis necessárias ao INE para efeitos do estudo do novo modelo censitário.

O conteúdo da proposta está condicionado pela Deliberação da CNPD nº 929/2014, de 11.06.2014.

2.7.2 Indicadores de qualidade

A avaliação da qualidade do conteúdo dos dados do IMI recebidos no INE envolve duas componentes: cobertura da base de dados e qualidade intrínseca a cada variável.

Cobertura:

Verificam-se algumas dificuldades na justificação das diferenças encontradas entre os dois “tipos de dados” recebidos, um em ritmo anual, outro em ritmo mensal. Considera-se como de maior qualidade a base mensal, no quadro da qual se contabilizam (outubro de 2014) cerca de **6 milhões** de matrizes. Este valor é coerente com o nº de alojamentos familiares apurados nos Censos 2011: cerca de **5,9 milhões**. É ainda necessário verificar a razão pela qual não se localizam, na referida versão da BD do IMI, alguns registos correspondentes a alojamentos específicos, utilizados para testar a cobertura do ficheiro.

Qualidade das variáveis:

No que respeita aos atributos de identificação, a análise que já foi possível efetuar aponta para que os mesmos sejam de boa qualidade e estejam preenchidos em 100% dos casos. Não foi possível verificar (por inexistente na versão disponível no INE) a qualidade do único identificador numérico utilizado, o NIF do titular. Porém, dada a natureza da fonte, não existem dúvidas neste domínio.

O mesmo pode ser afirmado relativamente a outras variáveis de identificação e de contato, as quais estão 100% preenchidas: Nome do titular, endereço, Código Postal (CP7), nº de matriz, coordenadas geográficas.

Relativamente aos endereços, detetam-se deficiências nos ficheiros enviados ao INE. Interessará aceder também às versões de endereço utilizadas pela AT no contato com os contribuintes, as quais têm maior qualidade.

Também derivado da natureza da fonte e associado ao objetivo de cobrança do imposto, é de 100% a taxa de preenchimento das variáveis de caracterização de

cada imóvel, dado que estão associadas ao cálculo do valor tributável. Porém, detetam-se problemas de qualidade e de normalização no preenchimento de algumas variáveis, para as quais importa estabelecer procedimentos de correção / melhoria. Estas dificuldades situam-se, essencialmente, nas variáveis “Nº de pisos” e “Tipologia / Nº de divisões”.

Algumas questões relativas à cobertura e à qualidade de algumas variáveis deverão ser aprofundadas em diálogo direto com a entidade gestora desta fonte administrativa (AT). Inclui-se neste âmbito a verificação de aparentes problemas na cobertura (embora de pequena dimensão) dos períodos associados à entrada em vigor do IMI (2003), assim como do início de vigência do protocolo com o INE (2005).

2.7.3 Identificação das variáveis que podem ser usadas para fins censitários

Nos Censos 2011, foram recolhidas 34 variáveis relacionadas com a habitação, 11 das quais obrigatórias por regulamento comunitário.

O quadro seguinte identifica as **12** variáveis censitárias, relacionadas com a área da habitação, que podem ser obtidas através do **IMI**:

Quadro 19 – Variáveis possíveis de obter através do IMI

Unidade Estatística	Variáveis Censos 2011 (Habitação), possíveis de obter via IMI
Edifício	Tipo de edifício (*)
Edifício	Nº de alojamentos (*)
Edifício	Época de construção (*)
Edifício	Tipo de utilização
Edifício	Nº de pisos
Edifício	Elevador
Alojamento	Água (*)
Alojamento	Retrete (*)

Alojamento	Área útil ou N° de divisões (*)
Alojamento	Esgotos
Alojamento	Estacionamento
Alojamento	Entidade proprietária

(*) Variáveis obrigatórias, por regulamento comunitário

Algumas variáveis têm relação direta com os elementos constantes do IMI, outras devem ser obtidas de forma derivada.

Não estão ainda criadas condições para uma análise detalhada, a nível de microdado, do conteúdo do IMI e respetiva comparação com os Censos 2011. Ainda assim, é possível antever que a fonte IMI apenas terá maior utilidade relativamente aos edifícios e alojamentos concluídos após o momento censitário (março de 2011). No que se refere ao stock de edifícios e alojamentos inventariado nos Censos 2011, o IMI poderá ser utilizado quando através dele se registarem atualizações posteriores a 03.2011.

2.8 Sistema de Indicadores de Operações Urbanísticas (SIOU)

2.8.1 Breve descrição

O SIOU tem como enquadramento legal mais recente a Portaria n.º 235/2013, de 24 de julho. Através desta legislação, é estabelecido o circuito de informação que, com origem nas câmaras municipais, faz chegar ao INE elementos estatísticos relativos a operações urbanísticas. A atual legislação substitui normas anteriores, que se considerou necessário atualizar. De qualquer modo, a parte central do projeto SIOU (licenciamento e conclusão de edifícios), já existe no INE há muitos anos, com outras designações.

O SIOU é composto por um conjunto de 6 elementos:

- Inquérito às Operações de Loteamento Urbano (Q1)
- Inquérito aos Trabalhos de Remodelação de Terrenos (Q2)

- Inquérito aos Projetos de Obras de Edificação e Demolição de Edifícios (Q3)
- Inquérito à Utilização de Obras Concluídas (Q4)
- Inquérito à Conclusão de Obras (Q5)
- Inquérito às Alterações de Utilização dos Edifícios (Q6)

O conteúdo do Q1 e Q2 não tem relevância para o estudo de viabilidade em curso. O elemento principal a considerar, em termos de variáveis é o Q3, em cujo conteúdo se suportam o Q4, Q5 e Q6.

Toda a informação constante dos questionários Q1 a Q4 e Q6 é comunicada mensalmente ao INE pelas câmaras municipais, em suporte digital. O Inquérito à Conclusão de Obras (Q5) envolve o contato com o promotor da obra quando, relativamente a uma determinada obra licenciada, o INE não recebeu (no prazo previsto para a sua conclusão) informação da autarquia (Q4) sobre a correspondente licença de utilização.

Na última revisão do SIOU, em vigor desde janeiro de 2013, foram consideradas algumas questões relacionadas com a atualização do FNA. Realça-se a georreferenciação dos edifícios licenciados e concluídos, assim como a identificação e caracterização (tipologia e área) dos fogos de cada edifício. De igual modo se procedeu à calendarização das tarefas associadas à recuperação desta (nova) informação, com recuo até ao momento censitário.

2.8.2 Indicadores de qualidade

Os dados do SIOU são validados, na fonte (CM) e no INE, pelo que é boa a qualidade intrínseca a cada variável.

Em termos de cobertura, é garantida a abrangência integral de todas as construções licenciadas, ficando de fora eventuais construções clandestinas. O licenciamento da construção envolve os vários destinos das obras, designadamente a “habitação familiar” e as “convivências”. Nos documentos

relativos ao SIOU é também mencionada uma outra componente, há muito existente: as construções isentas de licenciamento. Os contatos já efetuados com o DEE ainda não permitiram clarificar os procedimentos atualmente em vigor, nem os números envolvidos nesta modalidade. Contudo, os valores em causa serão certamente de pequena dimensão.

Uma atenção especial deve ser dada aos efeitos da relação edifício concluído / edifício com licença de utilização. Uma forma de, no âmbito da metodologia do SIOU, se obviar à tardia (ou não comunicação ao INE) emissão de licenças de utilização (informação a receber através do Q4), é o recurso ao contato direto com o promotor das obras, quando está terminado o prazo da licença e nada mais se sabe sobre a obra em causa. Para este efeito, é realizado o inquérito Q5. Segundo o DEE (vide “Estatísticas da Construção e Habitação - 2013”), *“a informação relativa à conclusão de obras é obtida maioritariamente por inquéritos dirigidos aos seus promotores. Este método de recolha origina atrasos substanciais na obtenção da informação...”*, sendo efetuada a *“estimação do prazo efetivo de conclusão de uma obra a partir do seu prazo previsto”*.

Interessa promover, junto do DEE, a redução do nível de dependência da informação proveniente dos “requerentes de obras” dado que, apesar da baixa pressão construtiva dos últimos anos, em 2021 será necessário considerar, com rigor, o parque habitacional recenseado em 03.2011 ajustado pelas construções novas e demolições entretanto ocorridas.

Anota-se que, entre o 2º trimestre de 2011 e o 2º trimestre de 2014 (últimos dados disponíveis) foram concluídos cerca de **74 000** alojamentos, em construções novas. Entre 2011 e 2013 foram demolidos cerca de **3 000** edifícios. A manter-se o atual ritmo de construção (muito mais lento do que na década anterior, na qual se registou um acréscimo de 16%), é provável que em 2021 existam mais cerca de 220 000 alojamentos do que em 2011 (acrécimo de 4%).

2.8.3 Identificação das variáveis que podem ser usadas para fins censitários

Nos Censos 2011, foram recolhidas 34 variáveis relacionadas com a habitação, 11 das quais obrigatórias por regulamento comunitário.

O quadro seguinte identifica as **9** variáveis censitárias, relacionadas com a área da habitação, que podem ser obtidas através do **SIOU**:

Quadro 20 – Variáveis possíveis de obter através do SIOU

Unidade Estatística	Variáveis Censos 2011 (Habitação), possíveis de obter via SIOU
Edifício	Tipo de edifício (*)
Edifício	Nº de alojamentos (*)
Edifício	Época de construção (*)
Edifício	Tipo de utilização
Edifício	Nº de pisos
Alojamento	Água (*)
Alojamento	Área útil ou Nº de divisões (*)
Alojamento	Esgotos
Alojamento	Estacionamento

(*) Variáveis obrigatórias, por regulamento comunitário

No novo modelo censitário, a fonte SIOU terá grande utilidade relativamente aos edifícios e alojamentos concluídos após o momento censitário (março de 2011). Através desta fonte (classificável quer como administrativa, quer como inquérito), é viável atualizar o stock inventariado nos Censos 2011, permitindo uma boa gestão na cobertura (exaustividade) do parque habitacional. Em paralelo, o SIOU permite caracterizar esse parque relativamente a um nº significativo de variáveis (9, das quais 5 obrigatórias para a EU).

3. Avaliação do contributo das fontes administrativas - apresentação de resultados

Nos pontos anteriores, descreveu-se cada um dos 8 ficheiros administrativos já analisados (de um total de 11 relevantes para o projeto), do ponto de vista do âmbito, cobertura e qualidade.

Neste ponto são apresentados os resultados tendo em conta a identificação das áreas não cobertas por informação administrativa, assim como a informação que é possível utilizar e respetivas limitações. Pretende-se deste modo proporcionar um quadro de análise objetivo, ao qual se associa uma reflexão sobre hipóteses de cenários para os Censos 2021.

O contínuo aumento das variáveis observadas pelos Censos em Portugal, ao contrário do que tem vindo a ocorrer em termos internacionais, dificulta a gestão desta questão. No programa correspondente ao estudo de viabilidade em curso, está prevista a consulta dos utilizadores nacionais de dados censitários. Nesta fase, consideram-se como referência as variáveis incluídas nos Censos 2011, decorrentes da regulamentação comunitária, sem prejuízo da visão mais global de âmbito nacional.

No quadro seguinte apresenta-se a evolução do **nº de variáveis incluídas nos últimos 3 recenseamentos**:

Quadro 21 - Variáveis incluídas nos censos 1991-2011

Unidade estatística Nº de variáveis por fonte	Censos 1991	Censos 2001	Censos 2011
Edifício	6	15	16
Alojamento	13	15	18
Família/Indivíduo	32	37	43
Total	51	67	77

Verifica-se que, em três décadas, o nº de variáveis incluídas no conjunto dos questionários dos Censos aumentou cerca de **50%**, passando de 51 em 1991 para 77 em 2011. Esta é uma situação que, inevitavelmente, dificultará a posição do INE na auscultação dos utilizadores sobre a adoção de um novo modelo censitário.

Contudo, apenas **35%** (27) das variáveis recolhidas nos Censos 2011 são obrigatórias para transmissão ao Eurostat, como se verifica no quadro seguinte. Por outro lado, apenas **56%** (15) dessas variáveis obrigatórias (**19%** das recolhidas através dos Censos 2011) devem ser disponibilizadas para a UE a um nível geográfico inferior à região NUTS II, mais concretamente a Freguesia.

Quadro 22 – Síntese de variáveis obrigatórias (União Europeia)

Unidade estatística Nº de variáveis por fonte	Censos 2011	UE 2011		
		Total	NUTS 2	Freg
Edifício	16	2	0	2
Alojamento	18	8	5	3
Família/Indivíduo	43	17	7	10
Total	77	27	12	15

O conhecimento sobre a informação que pode ser utilizada a partir dos 11 ficheiros para substituir variáveis censitárias encontra-se relativamente estabilizado e não é espectável que, a partir de informação administrativa, novas variáveis censitárias possam ser cobertas. No seu conjunto, estes ficheiros permitem obter informação completa para **44%** (12) variáveis comunitárias e **16%** (8) das restantes variáveis recolhidas nos Censos 2011 (de utilização exclusivamente nacional). Para algumas outras variáveis, de âmbito comunitário e nacional, existe uma cobertura parcial nas fontes em causa. Os quadros seguintes ilustram bem o potencial e as restrições das fontes administrativas inventariadas.

Quadro 23 – Síntese de variáveis (comunitárias e nacionais), com e sem fonte administrativa

Unidade estatística Nº de variáveis por fonte	Censos 2011	UE c/ fonte administrativa		UE s/ fonte administrativa	Nacionais c/ fonte administrativa		Nacionais s/ fonte administrativa
		Cobertura total	Cobertura parcial		Cobertura total	Cobertura parcial	
Edifício	16	2	0	0	3	0	11
Alojamento	18	4	0	4	3	3	4
Família/Indivíduo	43	6	6	5	2	6	18
Total	77	12	6	9	8	9	33

A análise do quadro permite verificar que a possibilidade de obter informação administrativa não é uniforme, de acordo com as diferentes unidades estatísticas. **Por exemplo,** nenhuma das variáveis relativas à unidade estatística família pode ser obtida por informação administrativa.

Quadro 24 – Variáveis obrigatórias no âmbito do regulamento comunitário, segundo a disponibilidade em ficheiros administrativos

Disponibilidade da informação em ficheiros administrativos	Número de variáveis	Variáveis
Totalmente disponível	12	Local de residência
		Sexo
		Idade
		Estado civil
		Nacionalidade
		Naturalidade
		Água
		Área útil
		Instalações sanitárias
		Época de construção

Disponibilidade da informação em ficheiros administrativos	Número de variáveis	Variáveis
		Tipo de edifício
		Tipo de alojamento
Parcialmente disponível	6	Nível de ensino completo
		Condição perante a Atividade Económica
		Profissão
		Ramo de Atividade Económica
		Situação na profissão
		Local de trabalho
Não disponível	9	Eventual residência no estrangeiro e ano de chegado ao país
		Local de residência 1 ano antes
		Relações de parentesco
		Tipo de família
		Dimensão da família
		Forma de ocupação
		Banho ou duche
		Tipo de aquecimento
		Regime de ocupação
Total	27	

A leitura dos quadros seguintes identifica a grande sobreposição de fontes relativamente a uma parte da informação inventariada. Enquanto para muitas variáveis não existe qualquer fonte, outras, como é o caso das variáveis demográficas (data de nascimento, sexo, estado civil), estão presentes num número muito significativo de ficheiros.

A análise de redundância da informação é um fator positivo, uma vez que permite a avaliação da qualidade e consistência da informação e da fonte. Esta análise introduz maior certeza na associação de informação (variáveis) proveniente de diferentes ficheiros a um único indivíduo.

Os dois quadros seguintes permitem identificar o contributo dos 11 ficheiros administrativos inventariados no âmbito do estudo de viabilidade para os Censos 2021. Até ao momento foram estudados 8 desses ficheiros e (apenas) inventariada a informação com potencial de utilização nos 3 restantes.

No primeiro quadro incluem-se as variáveis obrigatórias por regulamento comunitário, enquanto no segundo quadro se apresentam as restantes variáveis recolhidas nos Censos 2011.

Quadro 25 – Variáveis (comunitárias), com e sem fonte administrativa

Unidade estatística	Variáveis / Modalidades obrigatórias (União Europeia)	Cobertura Total / Parcial	Nº fontes admin	Fonte administrativa										
				SIOU	IMI	IRS	BDIC	SEF	ME	SS	CGA	QP	IEFP	ADSE
Variáveis obrigatórias, com 1 ou + fontes administrativas (assegurando diferentes níveis de cobertura do universo)														
Edifício_Alojamento	Tipo de edifício	T (*)	2	X	X									
	Época de construção	T (*)	2	X	X									
	Tipo de alojamento (familiar clássico)	T (*)	2	X	X									
	Área útil / N° divisões (obrig 1 de 2)	T (*)	2	X	X									
	Água	T (*)	2	X	X									
	Instalações sanitárias	T (*)	1		X									
Família_Indivíduo	Sexo	T	9			X	X	X	X	X	X	X	X	X
	Idade	T	9			X	X	X	X	X	X	X	X	X
	Naturalidade (País, Município, Freguesia)	T	4			X	X	X		X				
	Nacionalidade (País)	T	6			X	X	X	X	X			X	
	Residência habitual	T (**)	5				X	X	X	X	X			
	Nível de ensino completo	P	3						X			X	X	
	Estado civil legal	T	5			X	X	X		X			X	X
	Local de trabalho (País, Município, Freguesia)	P	4							X	X	X		
	CPA-população empregada	P	3							X	X	X		
	CPA-estudantes	T	1						X					
	CPA-reformados	P	2							X	X			
	CPA-pop desemp procura 1º emprego	P	1										X	
	CPA-pop desemp procura novo emprego	P	2							X			X	
	Profissão / população empregada	P	2					X				X		
	Profissão / pop desemp proc novo emprego	P	1										X	
	Situação Prof-patrões / empregadores	P	1									X		
	Situação Prof-trab conta de outrem	P	3							X	X	X		
	Situação Prof-trab conta própria	P	1							X				
	Situação Prof-outros (fam não remun, etc.)	P	1									X		
	Actividade económica / pop empregada	P	4							X	X	X		X
	Activ econ / pop desemp proc novo emprego	P	1										X	
Variáveis obrigatórias, sem qualquer fonte administrativa														
Edifício_Alojamento	Forma de ocupação		0											
	Regime de ocupação		0											
	Tipo de aquecimento		0											
	Banho ou duche		0											
Família_Indivíduo	Dimensão da família		0											
	Relação de parentesco		0											
	Tipo de família		0											
	Residência 1 ano antes		0											
	Residência fora de Portugal (alg vez / ano)		0											
	Situação Prof / pop desemp proc novo emprego		0											
	CPA-domésticos e outros		0											

Notas:

(*): Dadas as características das unidades estatísticas “edifício” e “alojamento”, em termos de estabilidade, assume-se que os Censos 2011 constituem uma “fonte administrativa” para o parque habitacional construído até março de 2011

(**): A variável “Residência habitual” não consta diretamente na principal fonte relacionada com a “População”, a BDIC. Está em estudo a sua estimativa por cruzamento dos dados dos Censos 2011 com informação proveniente de várias fontes (BDIC, SEF, Seg,Social, CGA, Educação).

Quadro 26 – Variáveis (nacionais), com e sem fonte administrativa

Unidade estatística	Variáveis nacionais (Censos 2011, não obrigatórias para a União Europeia)	Cobertura Total / Parcial	Nº fontes admin	Fonte administrativa										
				SIOU	IMI	IRS	BDIC	SEF	ME	SS	CGA	QP	IEFP	ADSE
Variáveis nacionais, com 1 ou + fontes administrativas (assegurando diferentes níveis de cobertura do universo)														
Edifício_Alojamento	Nº de pisos	T (*)	2	X	X									
	Elevador	T (*)	1		X									
	Tipo de utilização	T (*)	2	X	X									
	Esgotos	T (*)	2	X	X									
	Estacionamento	T (*)	2	X	X									
	Entidade proprietária	P	1		X									
	Encargos de aquisição	P	1			X								
	Escalaão de renda	P	1			X								
Família_Indivíduo	Companheiro em união de facto	P	3			X				X				X
	Frequência de ensino	T	1						X					
	Ano de escolaridade	T	1						X					
	Fonte de rendimento (trabalho, reforma, etc)	P	1			X								
	Procura emprego - tempo	P	1									X		
	Disponibilidade p/ trabalhar	P	1									X		
	Nº horas trabalhadas	P	1									X		
Nº pessoas da empresa	P	1									X			
Variáveis nacionais, sem qualquer fonte administrativa														
Edifício_Alojamento	Material de revestimento exterior		0											
	Estrutura de construção		0											
	Tipo de cobertura		0											
	Configuração R/C		0											
	Edifício isolado / alto		0											
	Edifício gaveto / geminado		0											
	Edifício + ou - alto que adjacentes		0											
	Necessidade de reparações		0											
	Acessibilidade da entrada		0											
	Acessibilidade até entrada no alojamento		0											
	Recolha de resíduos sólidos urbanos		0											
	Ar condicionado		0											
	Fonte de energia		0											
	Tipo de contrato de arrendamento		0											
	Época do contrato		0											
Família_Indivíduo	Presente momento censitário		0											
	Residência 5 anos antes		0											
	Resid fora de Portugal (país prov)		0											
	Curso superior		0											
	Alfabetismo		0											
	Tempo p/ local trabalho / estudo		0											
	Meio de transporte 1 p/ local trabalho / estudo		0											
	Meio de transporte 2 p/ local trabalho / estudo		0											
	Trabalhou alguma vez		0											
	Tarefas na profissão		0											
	Bens e serviços produzidos		0											
	Dificuldade ver		0											
	Dificuldade ouvir		0											
	Dificuldade andar		0											
	Dificuldade memória		0											
	Dificuldade banho / vestir		0											
	Dificuldade compreender		0											
Religião		0												

(*):Dadas as características das unidades estatísticas “edifício” e “alojamento”, em termos de estabilidade, assume-se que os Censos 2011constituem uma “fonte administrativa” para o parque habitacional construído até março de 2011.

Como se referiu antes, para uma parte das variáveis relativamente às quais existem fontes administrativas, verifica-se um problema de **cobertura** relativamente ao universo que se pretende quantificar ou caracterizar. No quadro seguinte sintetiza-se a análise realizada até à data do presente relatório.

Quadro 27 – Síntese da análise de cobertura relativa ao universo de algumas variáveis existentes em fontes administrativas

Unidade estatística	Variável	Nº fontes admin	Problemas de cobertura em variáveis com 1 ou + fontes administrativas	% não coberta da variável / modalidade (estimativa)
Variáveis obrigatórias (União Europeia)				
Família_Indivíduo	Nível de ensino completo	3	Não há informação para o total da população, apenas para os alunos de um determinado ano. A reconstituição do stock está fora do alcance do projeto.	76%
	Local de trabalho (País, Município, Freguesia)	4	Em falta os dados para os trabalhadores por conta própria e os sem qualquer regime de segurança social.	8%
	CPA-população empregada	3	Em falta os dados para outros regimes de segurança social, para além da SS e CGA. Em falta os indivíduos sem qualquer regime.	3%
	CPA-reformados	2	Em falta os dados para outros regimes de segurança social, para além da SS e CGA. Em falta os indivíduos sem qualquer regime nacional.	1%
	CPA-pop desemp procura 1º	1	Estudo de cobertura dos dados do IEFP ainda não efetuado.	?
	CPA-pop desemp procura novo emprego	2	Estudo de cobertura dos dados do IEFP ainda não efetuado.	?
	Profissão / população empregada	2	Trabalhadores da Administração Pública e trabalhadores por conta própria, sem fonte.	21%
	Profissão / pop desemp proc novo emprego	1	Estudo de cobertura dos dados do IEFP ainda não efetuado.	?
	Situação Prof-patrões / empregadores	1	Em falta os dados para outros regimes de segurança social, para além da SS e CGA. Em falta os indivíduos sem qualquer regime.	1%
	Situação Prof-trab conta de outrem	3	Em falta os dados para outros regimes de segurança social, para além da SS e CGA. Em falta os indivíduos sem qualquer regime.	2%
	Situação Prof-trab conta própria	1	Em falta os dados para outros regimes de segurança social, para além da SS e CGA. Em falta os indivíduos sem qualquer regime.	2%
	Situação Prof-outros (fam não remun, etc.)	1	Estudo de cobertura ainda não efetuado.	?
	Actividade económica / pop empregada	4	Em falta os dados para os trabalhadores por conta própria e os sem qualquer regime de segurança social. Dúvidas sobre o nível de desagregação da CAE, possível de obter para os trabalhadores da Administração Pública.	?
	Activ econ / pop desemp proc novo emprego	1	Estudo de cobertura dos dados do IEFP ainda não efetuado.	?
Variáveis nacionais				
Edifício_Alojamento	Entidade proprietária	1	Estudo de classificação de NIF (no IMI) ainda não efetuado.	?
	Encargos de aquisição	1	Estudo de cobertura do IRS ainda não efetuado.	?
	Escalão de renda	1	Estudo de cobertura do IRS ainda não efetuado.	?
Família_Indivíduo	Companheiro em união de facto	3	Estudo de cobertura do IRS ainda não efetuado.	?
	Fonte de rendimento (trabalho, reforma, etc)	1	Estudo de cobertura do IRS ainda não efetuado.	?
	Procura emprego - tempo	1	Estudo de cobertura dos dados do IEFP ainda não efetuado.	?
	Disponibilidade p/ trabalhar	1	Estudo de cobertura dos dados do IEFP ainda não efetuado.	?
	Nº horas trabalhadas	1	Trabalhadores da Administração Pública sem fonte.	15%
	Nº pessoas da empresa	1	Trabalhadores da Administração Pública sem fonte.	15%

4. Criação de uma Base de População Residente (BPR) a partir de registos administrativos

4.1 Ligação entre as várias bases de dados

Nos pontos anteriores deste relatório, foram caracterizados, de forma isolada, os diferentes ficheiros administrativos já analisados. Avaliou-se a sua qualidade e foram identificadas as variáveis com potencial de utilização para fins censitários.

Neste ponto é abordada a integração dos diferentes ficheiros administrativos, condição essencial para a constituição de uma base de população.

A constituição de uma base de dados com as características da população residente implica obrigatoriamente a ligação entre os registos provenientes dos diferentes ficheiros.

As ligações entre chaves numéricas são as mais fáceis de concretizar e conferem maior grau de certeza e qualidade ao processo de ligação entre ficheiros.

Em Portugal, a ausência de um número único do cidadão torna mais complexa a ligação entre os diferentes ficheiros administrativos.

O quadro abaixo resume a situação dos ficheiros analisados quanto ao preenchimento dos identificadores numéricos.

Quadro 28 – Bases de dados e identificadores numéricos

Variáveis / Identificadores	BD's avaliadas (1)							BD's não avaliadas (2)		
	BDIC (2010)	SEF (2011)	SS (2013)	CGA (2008)	QP (2011)	Educação (2011_12)	IMI	IRS	IEFP	ADSE
Nº de registos	11,2 M	434 mil	7,1 M	1,2 M	2,7 M	1,9 M	5,7 M	x	x	x
Existência de identificador e respetiva % de preenchimento:										
NIC_N.º Identificação Civil	100%	-	76%	72%	-	83%	-	-	x%	x%
NIF_N.º Identificação Fiscal	-	48%	98%	48%	-	-	-	x%	x%	x%
NISS_N.º Identificação Segurança Social	-	39%	100%	-	100%	-	-	-	x%	-
NIF_N.º Identificação Fiscal do proprietário	-	-	-	-	-	-	x%	-	-	-
NIF_N.º Identificação Fiscal da empresa	-	-	-	-	100%	-	-	-	-	-

- (1) *BD's analisadas com diferentes níveis de profundidade e de atualidade. Na maioria dos casos, apenas houve acesso a 1 ano de dados, nuns casos bastante antigos (CGA-2008), noutros casos com mais atualidade (SS-2013)*
- (2) *BD's não analisadas, por não ter ainda sido possível o acesso às mesmas*

Uma vez que a ligação entre ficheiros só parcialmente poderá ser concretizada através de identificadores numéricos, o desenvolvimento de chaves de *matching* tem vindo a ser testado de acordo com o seguinte esquema e ordem de prioridades.

- **Matching 1** – Ligações através de campos numéricos: NIC, NIF, NISS, etc.
- **Matching 2** – Ligações através de campos não numéricos, utilizando o nome dos indivíduos conjugado com outros campos.
Exemplo: Nome próprio (primeiros 3 caracteres do 1º nome) + Apelido (últimos 3 caracteres do último nome) + Data de nascimento (AAAAMMDD) + Sexo (1-Masculino; 2-Feminino).
- **Matching 3** – Ligações através de campos não numéricos utilizando o campo morada conjugado com outros campos (não testado, mas conhecendo-se as dificuldades deste tipo de matchings).
- **Outros...**

4.2 Metodologia para a construção da Base de População Residente

O esquema apresentado ilustra as diferentes fases até à constituição da Base de População Residente.

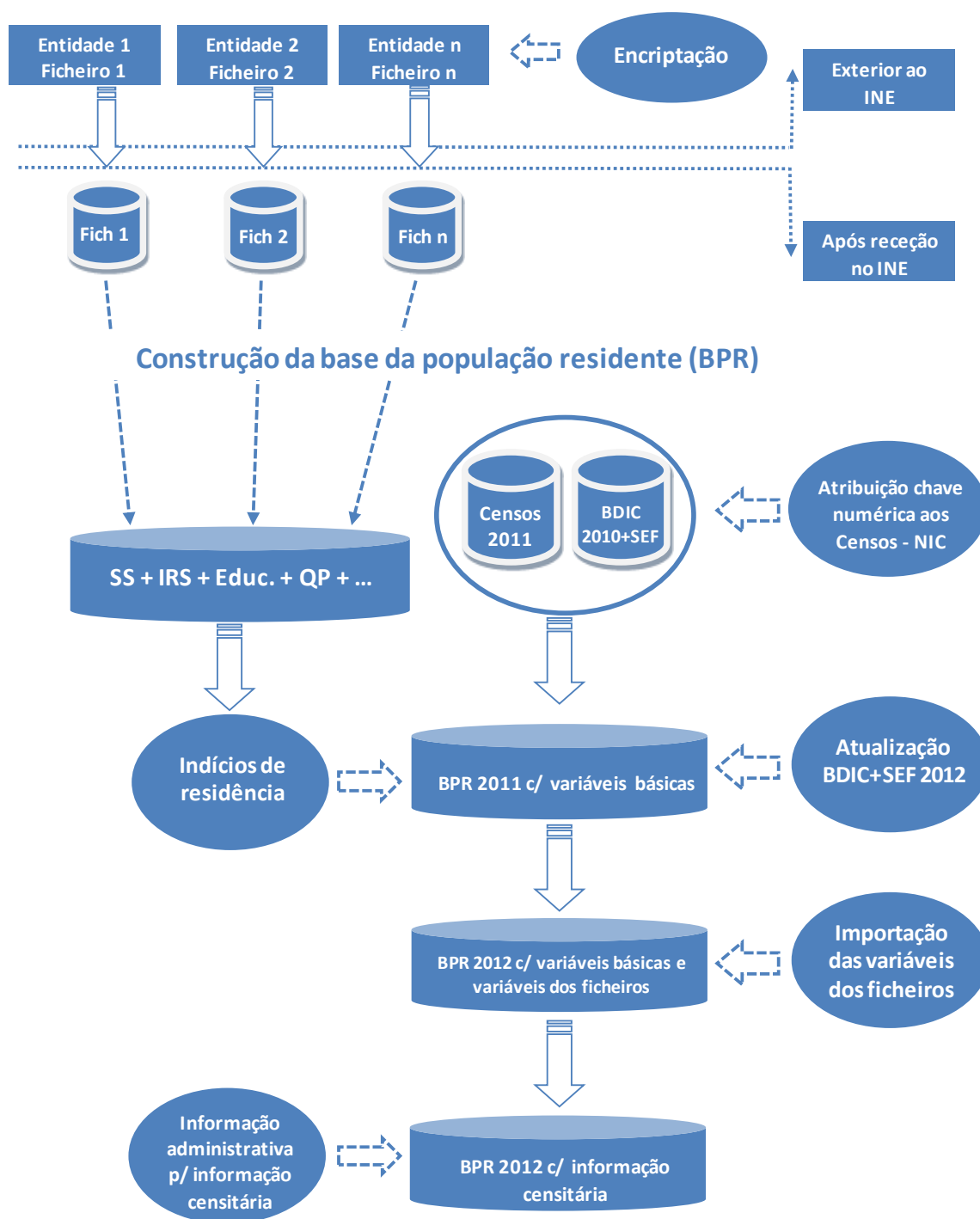
O processo inicia-se com a preparação e envio dos ficheiros por parte das entidades responsáveis. Cada entidade deve assegurar o cumprimento das regras de salvaguarda da confidencialidade dos registos individuais, procedendo à encriptação dos identificadores numéricos e abreviando o respetivo nome dos titulares, de acordo com a Deliberação da CNPD nº 929/2014, de 11 de junho.

O INE apoiou este processo, facultando a respetiva aplicação de encriptação.

Segue-se uma segunda etapa de receção, armazenamento dos ficheiros, verificação da qualidade e aplicação de regras de normalização. Após concluída esta fase, os ficheiros estão em condições de poder ser utilizados e ligados entre si.

As etapas seguintes descrevem a aplicação da metodologia que visa, numa primeira fase, a constituição do ficheiro de população residente e posteriormente a caracterização dos indivíduos, através da incorporação de variáveis no modelo de dados.

Fig.8 – Encriptação, análise e pré-tratamento dos ficheiros e construção da BPR



A construção do modelo para a constituição da BPR (Base de população residente) parte dos indivíduos recenseados nos Censos 2011 (espinhal dorsal do modelo) e compreende as seguintes 4 etapas:

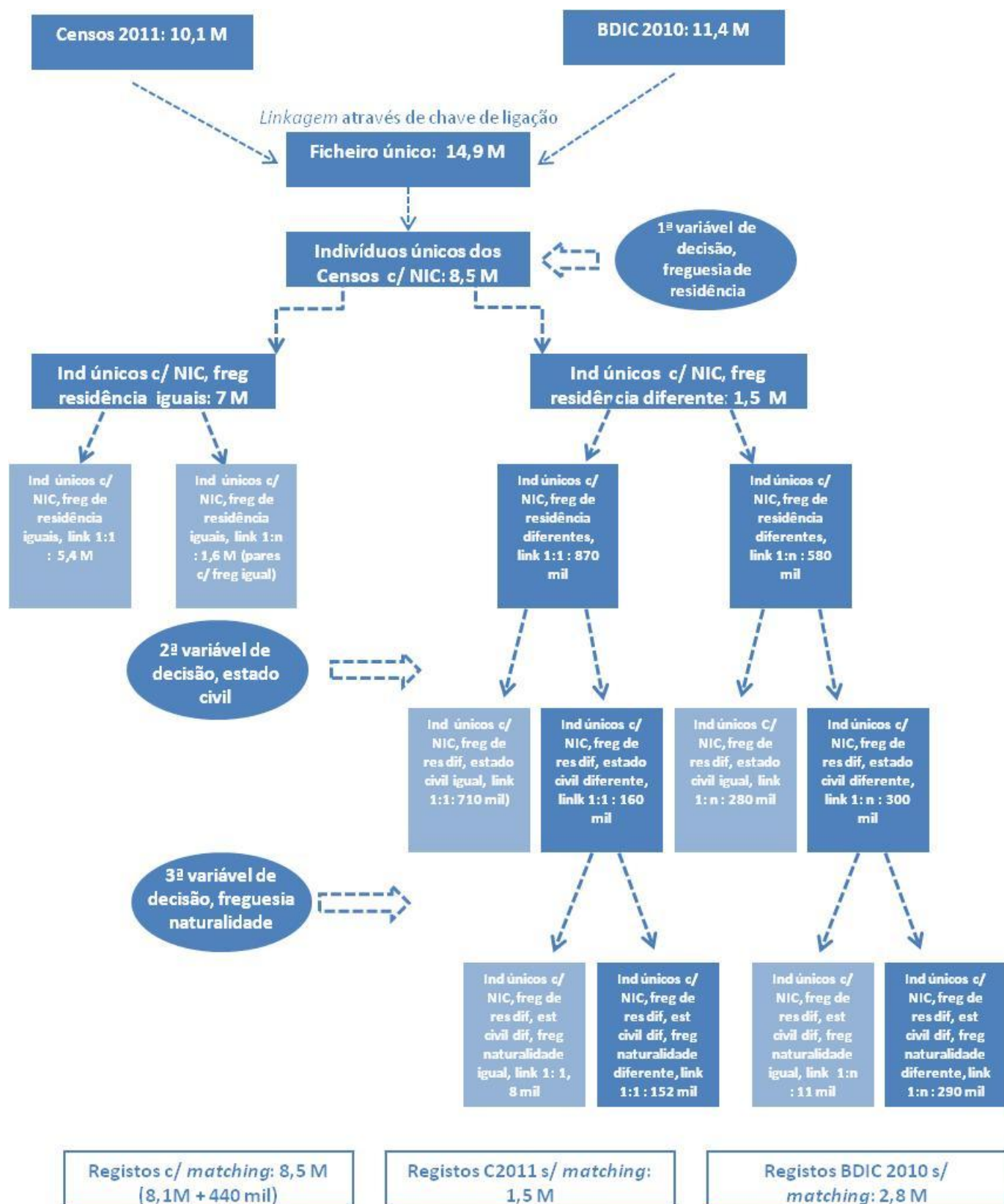
1ª Etapa: Colocação de NIC 's nos Censos 2011 e criação da BPR 2011

A colocação de identificadores numéricos (NIC) na Base dos Censos 2011 é essencial para permitir a ligação entre os Censos 2011 e os ficheiros administrativos.

Tendo em conta que os Censos 2011 não possuem nenhuma chave numérica, a ligação entre os Censos 2011 e a BDIC 2010, foi efetuada através de um algoritmo que combinou vários atributos - matching tipo 2 - descrito anteriormente.

O processo de matching foi efetuado por etapas de acordo com o esquema apresentado na página seguinte. O processo encontra-se descrito em anexo.

Fig. 9 - Esquema de colocação de NIC's nos Censos 2011 através de chaves não numéricas



Principais resultados e conclusões:

- De um total de 10,1 milhões de registos de indivíduos portugueses dos Censos 2011, 8,5 milhões (84%) efetuaram *matching* com a BDIC 2010 através da chave de ligação não numérica, o que permitiu a atribuição do respetivo NIC à base dos Censos 2011.
- Em 8,2 milhões de registos dos Censos 2011 (81% da população residente portuguesa) aos quais foi atribuído o NIC, para além de se ter verificado a coincidência nas variáveis que integram a chave de matching (data de nascimento, sexo e 3 primeiras letras do 1º nome, 3 últimas letras do último nome), existia pelo menos mais uma variável coincidente entre a BDIC 2010 e os Censos 2011 (Freguesia de residência ou Estado Civil ou Naturalidade).
- Cerca de 7 milhões de registos dos Censos 2011 (69% da população residente portuguesa) efetuaram matching através da chave de ligação e possuíam também a mesma freguesia de residência nos Censos 2011 e na BDIC 2010.
- Cerca de 1,2 milhões de registos dos Censos 2011 (11,9% da população residente portuguesa) efetuaram matching através da chave de ligação e possuíam ou o mesmo estado civil ou a mesma naturalidade, mas não tinham a mesma freguesia de residência.
- Em 400 mil registos dos Censos 2011 (3,9% da população residente portuguesa) que fizeram matching através da chave, não coincide a freguesia de residência nem o estado civil legal nem a naturalidade.
- Dos cerca de 1,5 milhões de registos censitários (de um total de 10,1 milhões) que não efetuaram matching com a BDIC, conseguiram ser identificados na SS e na Educação cerca de 240 mil. Destes, cerca de 185 mil (77%) possuíam NIC.
- O número de registos censitários (população portuguesa) relativamente aos quais não se conseguiu atribuir um identificador numérico foi de cerca de **1,3 milhões**.
- Tendo em conta que para a realização deste exercício não foi utilizado nenhum software específico de matching, os resultados obtidos através da

chave construída, são muito positivos e permitem antever que existe um grande potencial de melhoria.

- A análise de qualidade dos resultados foi efetuada comparando, para o mesmo NIC a informação registada em três bases de dados, (Censos 2011, BDIC e SS).
- Para os registos com o mesmo NIC nos Censos, na BDIC e na Segurança Social (4,1 milhões), concluiu-se que em 95% dos casos (3,9 milhões de registos) se tratava do mesmo indivíduo, uma vez que possuía nas três BD a mesma data nascimento, o mesmo sexo e o mesmo nome completo.
- Foram identificados 170 mil indivíduos com o mesmo NIC, mas cujo nome completo diferia na base de dados da SS. Uma parte significativa de casos refere-se contudo à mesma pessoa, uma vez que a diferença no nome se relaciona com abreviaturas (maioritariamente quando o 1º nome é Maria, na SS é M. e nos restantes ficheiros Maria).
- É necessário efetuar um estudo de qualidade mais aprofundado dos resultados do matching obtido, uma vez que é fundamental que a atribuição do NIC ao registo do ficheiro dos Censos 2011, seja o mais exato possível, já que esse registo passará a ligar-se através desse identificador com os vários ficheiros administrativos. Também neste caso, os resultados obtidos neste exercício parecem conferir um grau de certeza elevado a pelo menos 7 milhões de registos (69% de indivíduos dos Censos 2011 de nacionalidade portuguesa).
- Espera-se que o recurso a técnicas de matching e a utilização de softwares apropriados, a par da melhoria da cobertura do ficheiro da BDIC, permita taxas de colocação de NIC na base dos Censos 2011 próximos dos 100%.

2ª Etapa: Processo de atualização da BPR 2011

O modelo de atualização anual da BPR baseia-se nas alterações verificadas na BDIC e no SEF. A análise das imagens dos respetivos ficheiros, em dois anos consecutivos, permite identificar as alterações, as quais possibilitam atualizar a informação da BPR. Por esta via, ganha-se eficiência, face à utilização de toda a

informação da BDIC ou do SEF e assegura-se uma estrutura geograficamente coerente, baseada nos Censos 2011.

As alterações verificadas de um ano para o outro nos ficheiros quer da BDIC quer do SEF, podem ser de dois tipos: Entrada /saída de registos (pessoas); alteração das variáveis associadas aos indivíduos (nome, estado civil, morada de residência, etc.).

Tendo em conta o tipo de alteração (variável) serão definidas regras de decisão para incorporação na BPR. O caso concreto da morada pode levantar alguns problemas, tendo em conta o conceito estatístico de população residente e o conceito de morada da BDIC (morada para contacto) que permite o apuramento de população registada.

A melhor afinação e detalhe metodológico deste processo, só poderá ser explicitada após analisar vários anos da BDIC, o que não foi possível até agora, uma vez que apenas dispomos da BDIC 2010.

3ª Etapa: Avaliação da população na BPR e confirmação de residência

Após atualização da BPR com base na BDIC e no SEF é necessário garantir que todos os indivíduos residentes em Portugal constam na base, e que por outro lado, não há indivíduos que constam da BPR mas que efetivamente não vivem em Portugal.

A avaliação da população e respetiva análise de residência baseia-se num algoritmo que tem por base indícios de residência, através da identificação do indivíduo num ou mais ficheiros. A análise dos resultados dá origem a uma matriz output (matriz de indícios de residência) à qual serão aplicadas regras de decisão relativamente a cada registo

Fig 10 - Matriz de indícios de residência

Indivíduos	Ficheiros administrativos				Regras de decisão
	SS	IRS	Educação		
Ind A			√		1
Ind B	√	√			2
Ind C		√	√		3
.....	√	√	√		
....					
Ind X			√		4
Ind Y	√	√			5
....					

O quadro abaixo permite identificar que a maioria da população residente em Portugal, tem pelo menos uma ligação com o Estado, uma vez que está presente em um ou mais ficheiros administrativos. Se somarmos o número de pessoas presentes na segurança social com o número de alunos que frequenta o ensino (ficheiros cujo grau de interseção é praticamente nulo) ficam cobertas mais de 9 milhões de pessoas, ou seja, menos 1 milhão relativamente ao total da população apurada pelos Censos 2011. Este indicador tem tendência a aumentar, através da integração de ficheiros como o IRS que abrangem franjas da população que podem não estar cobertas pelos ficheiros já estudados, assim como o ficheiro do IEF. Em síntese, podemos concluir que os ficheiros administrativos identificados contêm quase a totalidade da população residente em Portugal à data dos Censos 2011, sendo por isso uma fonte importante para validar a metodologia relativa à contagem da população residente.

Quadro 29 - Articulação entre os diferentes ficheiros através do NIC

	Únicos	SS	EDC.	CGA	SS-ED.	SS-CGA	ED-CGA	SS-ED-CGA	Total
Total NIC	8 362 725	6 384 864	1 738 769	770 087	597 088	291 817	39 450	21 868	8 893 720
NIC	6 758 490	5 144 434	1 451 927	770 087					7 366 448
Chave 2	1 604 235	1 240 430	286 842						1 527 272
S/matching		739 369	160 562	415 286					1 315 217
Nº Registos		7 124 233	1 899 331	1 185 373					10 208 937

4ª Etapa: Incorporação de variáveis na BPR a partir dos ficheiros administrativos

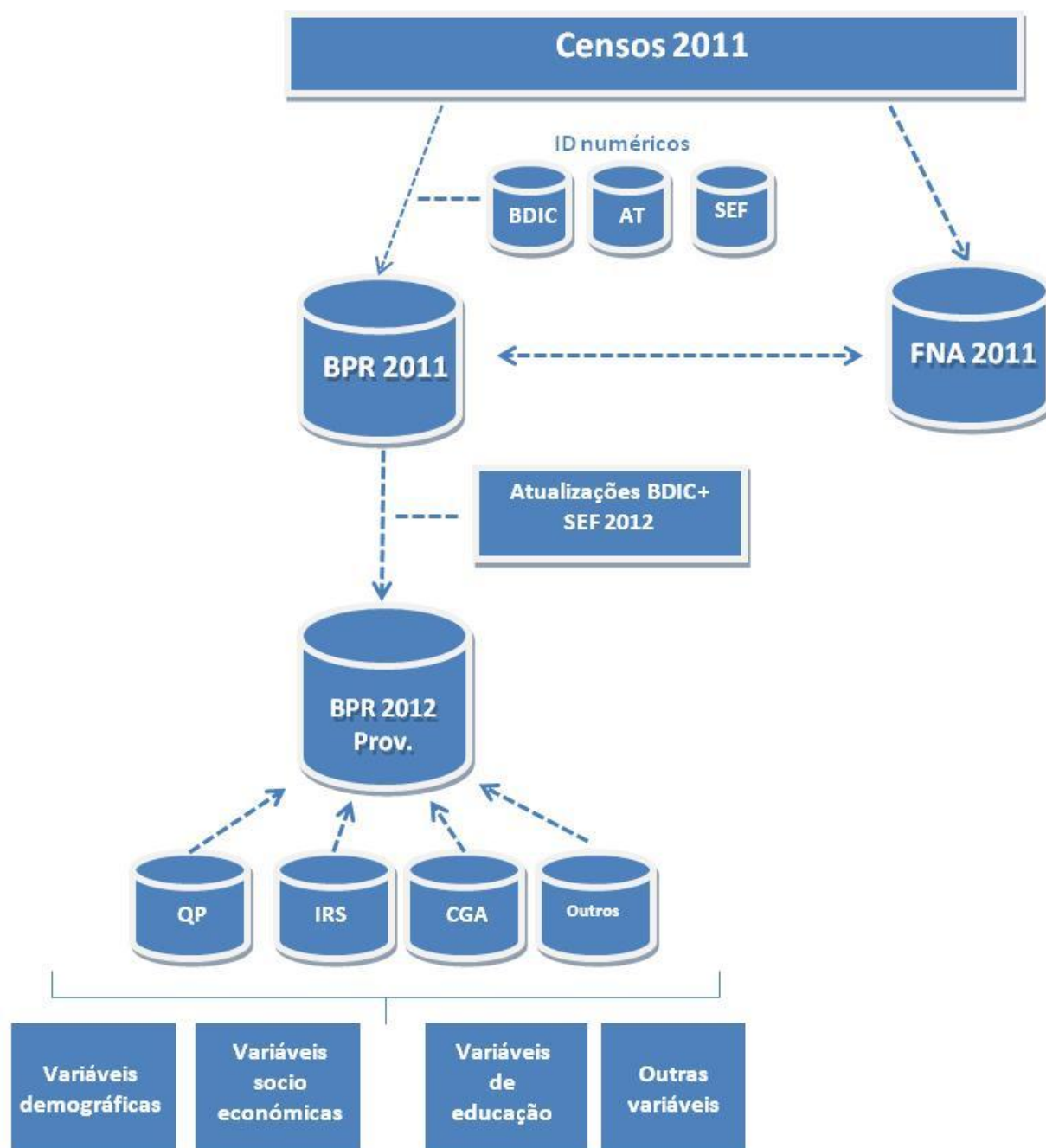
A inclusão de informação a partir dos ficheiros administrativos, sofre um processo prévio de normalização e tratamento que visa a transformação de informação administrativa em informação estatística.

Nesta etapa vai ser necessário identificar as variáveis e os respetivos ficheiros de proveniência, uma vez que há variáveis que são simultaneamente disponibilizadas em várias fontes. Como regra, será atribuída prioridade à informação proveniente das entidades responsáveis. Por exemplo: no caso das características demográficas, a prioridade é atribuída à BDIC, enquanto que relativamente às características do emprego e mercado de trabalho, a prioridade é atribuída à SS. O que significa que para um mesmo registo são várias as fontes que o alimentam, neste exemplo uma parte da informação é recolhida na base BDIC e outra na SS (mesmo que da SS constem dados demográficos).

A definição de regras de decisão para esta etapa, só poderão ser descritas após a implementação do modelo e análise dos resultados e das situações concretas.

A figura seguinte ilustra o esquema de constituição e atualização da BPR, de acordo com as etapas anteriormente descritas.

Fig. 11 – Esquematização da metodologia para a constituição da BPR



5. Avaliação do Programa de Trabalho 2013-2017 e conclusões

Durante o ano de **2014**, foram concretizadas todas as ações previstas no Programa de Trabalho 2013-2017:

- Avaliação dos modelos censitários utilizados noutros países, tendo em conta a sua adequação à realidade portuguesa;
- Criação de condições legais para acesso aos ficheiros administrativos, através da Deliberação da CNPD nº 929/2014, de 11 de junho;
- Articulação institucional com as diferentes entidades responsáveis pelos ficheiros administrativos; implementação da Deliberação;
- Análise dos ficheiros administrativos e o seu contributo para os Censos 2021;
- Definição da metodologia e modelo de dados, para constituição da Base de População Residente.

As principais conclusões do trabalho desenvolvido são:

Conclusão 1

Desenho do modelo para os Censos 2021 passa pela realização de um inquérito

Os trabalhos desenvolvidos durante o ano de 2014 identificaram a **necessidade de o novo modelo incluir a realização de um inquérito censitário para recolha de informação não disponível em ficheiros administrativos**, como deriva dos quadros apresentados anteriormente.

Destacam-se como lacunas essenciais as temáticas da **família** e da **educação**, assim como para diversas variáveis relativas ao indivíduo para as quais os ficheiros administrativos não permitem uma resposta direta ou aproximada.

É ainda necessário aprofundar o estudo correspondente à componente da habitação, a qual envolve os edifícios e os alojamentos familiares. Esta constitui uma das áreas onde a informação administrativa tem um maior potencial para substituir a recolha através de inquérito.

O desenho de um modelo censitário mais eficiente, que combine a realização de um inquérito com a utilização de informação administrativa, na componente da habitação parece ser uma opção sustentada pelos resultados obtidos até ao momento. A informação administrativa pode substituir totalmente cerca de **60%** das variáveis obrigatórias a nível comunitário relativas ao edifício/alojamento.

Quadro 30 - Informação censitária que pode ser substituída por informação administrativa, por tipo de unidade estatística

	Total	Edifício/ Alojamento	Família	Indivíduo
Totalmente disponível	44% (12)	60% (6)	0%	43% (6)
Parcialmente disponível	22% (6)	0%	0%	43% (6)
Não disponível	33% (9)	40% (4)	100% (3)	14% (2)

Na presente fase dos trabalhos, são consideradas as variáveis obrigatórias por regulamento comunitário (**censo de base zero**). A eventual inclusão de variáveis adicionais resultará da discussão a fazer com utilizadores preferenciais e da análise a efetuar no âmbito do Conselho Superior de Estatística.

O desenvolvimento deste cenário permitirá reduzir a recolha de informação através de inquérito, com os consequentes ganhos (financeiros e carga estatística) que daí resultam. Poderão ser tomadas opções relativamente ao

parque habitacional já recenseado em 2011, diferentes daquelas a aplicar às novas construções, construídas entre 2011 e 2021.

No quadro seguinte apresenta-se uma **síntese / avaliação das alterações propostas ao modelo censitário**.

Quadro 31 - Síntese das alterações propostas ao modelo censitário

Fator	Censos 2011	Censos 2021 (cenário)
Utilização de ficheiro de alojamentos	N	S
Distribuição e recolha de questionários porta a porta	S	N
Distribuição e recolha postal de questionários/PIN	N	S
Recolha de questionários pela Internet	S	S
Apuramento e divulgação de parte das variáveis censitárias com base em fontes administrativas	N	S
Carga estatística para o respondente	+	-
Complexidade e volume associado ao tratamento dos questionários	+	-
Custos com a operação *	+	-

Nota: Estima-se que a alteração do modo de recolha, através da utilização de um ficheiro de alojamentos, possa permitir uma redução de 40% nos custos da operação.

Conclusão 2

Constituição de uma Base de Dados de População Residente, a partir de ficheiros administrativos

A Base de Dados da População Residente (BPR) é o repositório de informação com as características da população que podem ser atualizadas por via administrativa e de modo mais alargado incluir também a informação proveniente de inquéritos.

A BPR, é um passo importante e decisivo no quadro dos trabalhos dos Censos 2021, mas também das **estatísticas sociais e nos censos pós-2021**, visando o apuramento e divulgação anual de parte das variáveis censitárias. Este trabalho está fortemente ligado às novas necessidades comunitárias relacionadas com a qualidade estatística da aplicação do conceito de população residente, com recentes implicações na metodologia de votação no Conselho Europeu. Serão aprofundados os trabalhos que interligam as atividades do GC 2021 com o DES, o qual está a desenvolver o estudo de viabilidade previsto no Art.º 8 do Regulamento nº 1260/2013, de 20.11.2013, relativo às estatísticas demográficas europeias.

Em 2014 foram criadas as condições para o desenvolvimento da BPR, as quais estão suportadas em quatro pilares:

- Condições legais de acesso aos ficheiros administrativos: espera-se que durante janeiro de 2015 o INE tenha acesso aos 11 ficheiros identificados no estudo;
- Desenho da metodologia de atualização da população residente dos Censos 2011, através da BDIC e do SEF, mas também de ficheiros complementares que permitam confirmar a residência dos indivíduos com base em indícios;
- Estudo de chaves não numéricas para ligação entre ficheiros;
- Desenho da infraestrutura da BPR e do respetivo modelo de dados.

6. Próximos desenvolvimentos e calendarização

1. Operacionalização do modelo para a constituição da BPR e avaliação dos resultados;
2. Desenvolvimento de um programa de trabalho que incorpore as necessidades do novo modelo censitário nos trabalhos do Ficheiro Nacional de Alojamentos;
3. Teste de avaliação do modelo para os Censos 2021.

No final de 2015 (outubro), está prevista a realização de um pequeno teste de campo, para avaliação dos resultados obtidos e afinação da proposta de modelo, a testar em 2016 .

Em articulação com o DRI, serão observados todos os edifícios, alojamentos e indivíduos residentes, num número restrito de áreas (algumas freguesias completas). Será recolhida informação para um número reduzido de variáveis, apenas as necessárias para permitir um confronto com os dados então disponíveis no INE e testar as metodologias em desenvolvimento relativamente a:

População Residente → Número; caraterísticas demográficas.

Ficheiro Nacional de Alojamentos → Exaustividade; endereço; forma de ocupação.

Em paralelo com a preparação, execução e avaliação deste pequeno teste, **será iniciada em 2015 a preparação do teste (de grande dimensão) a realizar em 2016**. Da avaliação dos resultados deste teste resultará a proposta final para o novo modelo censitário, a aplicar em 2021.

Quadro 32 – Calendário 2015/2016

Censos 2021 / Estudo de viabilidade - Calendário 2015 / 2016		Data início	Data fim	UO's
1 - Base de Dados de População Residente (BPR), a partir de ficheiros administrativos				
1.1	Receção de dados administrativos	12.2014	01.2015	GC, DMSI
1.2	Normalização de ficheiros	11.2014	02.2015	GC, DMSI
1.3	Introdução de chaves numéricas na BD dos Censos 2011	10.2014	03.2015	GC, DMSI
1.4	Matching de ficheiros	03.2015	04.2015	GC, DMSI
1.5	Construção da infraestrutura da BPR	02.2015	03.2015	DMSI
1.6	Aplicação da metodologia para cálculo da população residente	04.2015	05.2015	GC, DMSI
2 - Preparação do modelo para os Censos 2021				
2.1 - Principais fases de nível geral				
2.1.1	Análise do relatório QUAR 2014	-	01.2015	CD, GC
2.1.2	Auscultação de peritos e utilizadores especializados	02.2015	03.2015	CD, GC, SCI
2.1.3	Decisão sobre o modelo para os Censos 2021	-	06.2015	CD
2.1.4	Análise do relatório do teste 2016	-	12.2016	CD, GC
2.2 - Constituição do ficheiro censitário 2021, a partir do FNA				
2.2.1	Definição de objetivos para o ficheiro censitário	01.2015	02.2015	GC
2.2.2	Avaliação do FNA no contexto dos Censos 2021	02.2015	03.2015	GC, DMSI
2.2.3	Definição da metodologia de constituição do ficheiro censitário	04.2015	06.2015	GC, DMSI
2.2.4	Constituição do ficheiro de endereços para o teste 2016	02.2016	02.2016	DMSI
2.2.5	Construção do modelo de dados para constituição do ficheiro censitário	07.2016	12.2016	GC, DMSI
2.2.6	Operacionalização da metodologia de constituição do ficheiro censitário	2017	2019	DMSI
2.3 - Teste (mini) 2015				
2.3.1	Documento metodológico	01.2015	02.2015	GC
2.3.2	Especificações informáticas	02.2015	03.2015	GC, DMSI
2.3.3	Recolha de dados	10.2015	10.2015	GC, DMSI, DRI
2.3.4	Relatório	-	12.2015	GC
2.4 - Teste 2016				
2.4.1	Documento metodológico	03.2015	06.2015	GC
2.4.2	Especificações informáticas	05.2015	06.2015	GC, DMSI
2.4.3	Impressão de questionários e PIN's	03.2016	04.2016	GC, DMSI
2.4.4	Expedição postal	04.2016	04.2016	DMSI, DAFF
2.4.5	Recolha de dados	04.2016	05.2016	GC, DMSI, DRI
2.4.6	Relatório	-	12.2016	GC

ANEXOS

Anexo 1 – Metodologia para colocação de identificador numérico (NIC) nos censos 2011

A colocação de identificadores numéricos nos registos censitários recorreu a uma técnica de matching complexa que comparou os dados da BDIC 2010 com os dados dos Censos 2011 e identificou registos comuns.

Este processo desenvolveu-se em várias fases:

- **A primeira fase**, um estudo detalhado sobre as fontes, identificação das variáveis e análise da qualidade dos dados individualmente, foi objeto de desenvolvimento anterior e os resultados foram apresentados em relatórios específicos.
- A fase seguinte, **a segunda**, passou por estabelecer uma estratégia de ligação entre as duas fontes para construir um ficheiro resultante do matching com registos comuns.

A ligação entre duas bases de dados pode ser feita através de um campo numérico (normalmente um identificador numérico) presente em ambos os ficheiros ou, em alternativa, efetuar a ligação através de campos não numéricos.

Atendendo a que a base de dados dos Censos 2011 não possui nenhum identificador numérico em comum com a BDIC 2010, foi necessário criar uma chave de ligação de campos não numéricos- quando não existe uma chave única de ligação cria-se uma através da utilização de variáveis presentes em ambas as fontes.

De acordo com vários autores, na escolha da chave de ligação só devem entrar variáveis que não se alteram ao longo do tempo. Também segundo vários autores, quanto maior é o número de variáveis que constituem a chave de ligação menor são os registos ligados mas maior é o grau de certeza, o que leva a que a definição da chave não seja imediata, tendo que existir vários ensaios para que o resultado caiba dentro dos parâmetros pretendidos (confiança no resultado, número de casos obtidos,...).

O GC 2021, baseado nestes pressupostos, desenvolveu vários estudos para a definição das variáveis a integrar a chave de ligação, tendo estabelecido uma que pareceu viável e que foi testada. A tarefa de construção/teste das chaves ainda não se encontra finalizada.

- Na **terceira fase**, eliminaram-se as repetições resultantes da combinação através da chave de ligação e decidiu-se sobre a constituição do “par”. Após efetuar a comparação dos Censos 2011 com a BDIC 2010 ao nível dos registos individuais utilizando uma chave de ligação construída, pode acontecer o mesmo registo (indivíduo) estar “repetido” várias vezes. Esta situação decorre da forma como o processo se desenvolve e das próprias características das fontes em causa que levam à obtenção de vários resultados:
 - Matching 1 para 1, quando a um registo censitário corresponde um registo da BDIC, sendo o resultado que oferece maior segurança.
 - Matching 1 para n, quando a um indivíduo dos censos corresponde simultaneamente vários registos da BDIC (nesta situação o mesmo indivíduo dos censos repete-se com NIC diferentes).
 - Matching n para 1, quando vários indivíduos dos censos correspondem a um único registo da BDIC (nesta situação vários indivíduos dos censos apresentam o mesmo NIC).

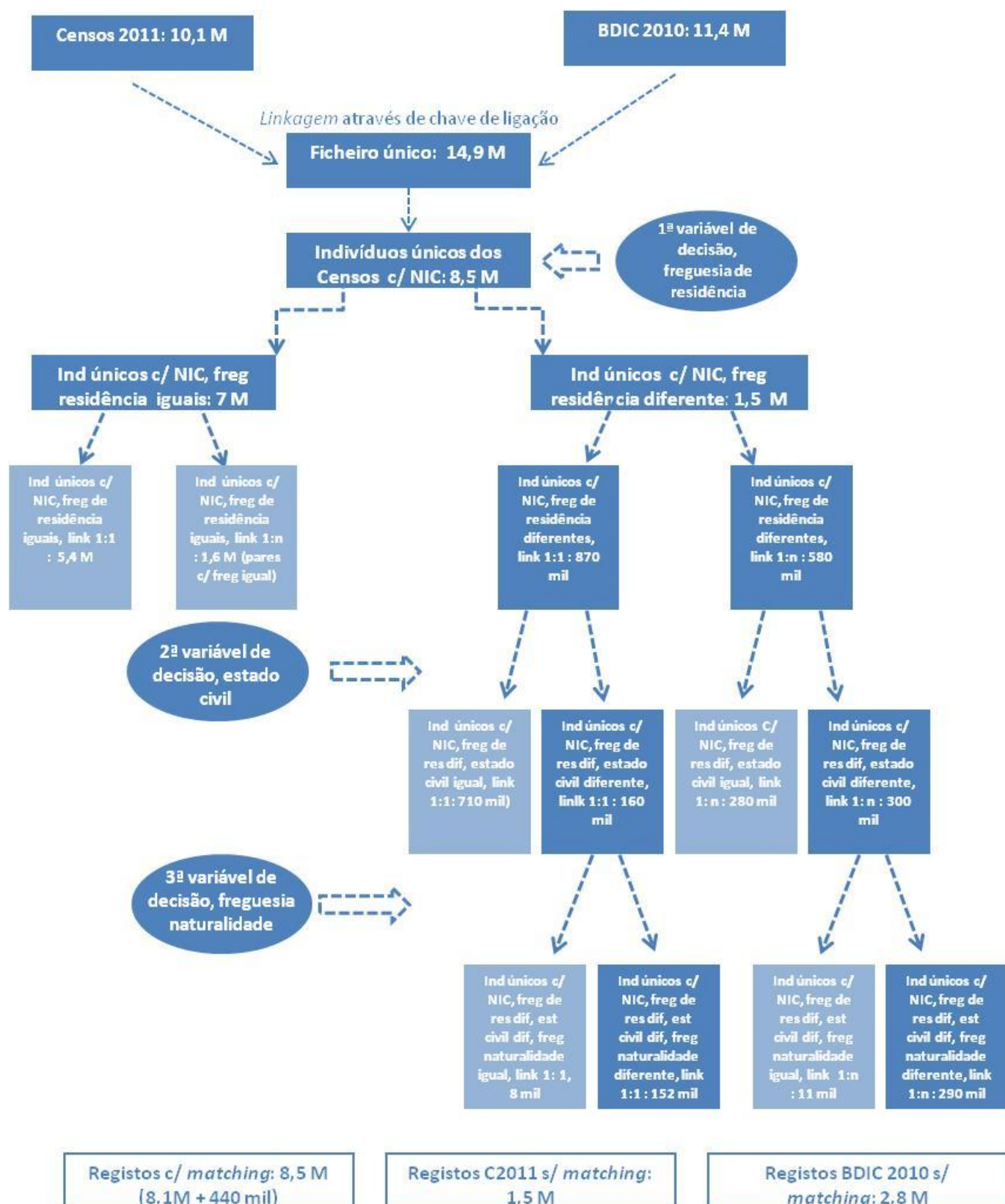
No 2º (1:n) e 3º(n:1) resultado de matching, há que decidir sobre o par a constituir (1:1), utilizando para isso as denominadas variáveis de decisão que são variáveis comuns aos dois ficheiros e determinam em que medida um par se refere ao mesmo indivíduo ou não.

- Na **quarta fase**, no sentido de obter alguns indicadores de qualidade, caracterizaram-se os resultados obtidos relativamente aos registos que efetuaram matching e aos registos que sobraram em cada fonte, e com o objetivo de conseguir colocar o máximo de NIC nos registos censitários procurou-se para estes últimos, o matching com outras fontes.

Assim, **para cada registo único dos Censos 2011 obtém-se o registo da BDIC 2010 que lhe corresponde, através de uma chave de ligação, utilizando outras variáveis que conferem determinado grau de certeza ao**

resultado. Após obter a correspondência (matching), é possível atribuir um NIC a um indivíduo proveniente da base de dados dos Censos 2011.

Em esquema, o processo é o seguinte:



O ficheiro utilizado dos Censos 2011 contém 10,1 milhões de registos porque se eliminaram os registos correspondentes à população de nacionalidade estrangeira: atendendo a que a BDIC é uma base de indivíduos de nacionalidade portuguesa que podem ou não ser residentes em Portugal, não é expectável o matching com indivíduos dos censos de nacionalidade estrangeira. Este grupo específico terá que ser comparado com o ficheiro proveniente do Serviço de Estrangeiros e Fronteiras (SEF).

O ficheiro utilizado relativo à BDIC 2010 contém 11,4 milhões de registos porque foram eliminados os registos com morada no estrangeiro, ou seja, portugueses que indicam residência no estrangeiro.

Para efetuar o matching entre os dois ficheiros através da ligação de variáveis não numéricas criou-se e testou-se uma **chave de ligação** com os Primeiros 3 caracteres do 1º Nome + Últimos 3 caracteres do último Nome + Data Nascimento (AAAAMMDD) + Sexo (1-Masculino; 2- Feminino).

Sempre que a chave de ligação era identificada em registos de ambos os ficheiros criou-se um ficheiro único, com cerca de 15 milhões de registos, no qual o mesmo indivíduo podia estar n vezes repetido, o que implicou ter que eliminar duplicados e determinar os pares.

Quadro 1 - Resumo dos resultados do ensaio utilizando a freguesia como variável de decisão

Nº Repetições do indivíduo dos censos	Ficheiro inicial(após Matching)	Indivíduos dos censos UNICOS	Freguesia da BDIC = FREGUESIA dos Censos 2011 (Diferença=0)	Freguesia da BDIC diferente FREGUESIA dos Censos 2011 (Indivíduos unicos)	Ficheiro final com chaves comuns matching (col 4+col5)	Registos a tratar (col3-col6)
col 1	col 2	col 3	col 4	col 5	col 6	col 7
1	6 460 711	6 460 711	5 395 477	869 312	6 264 789	195 922
2	1 933 574	966 787	792 092		792 092	174 695
3	1 117 698	372 566	301 229		301 229	71 337
4	797 228	199 307	160 473		160 473	38 834
5	618 330	123 666	99 228		99 228	24 438
6	501 768	83 628	67 168		67 168	16 460
7	404 348	57 764	46 482		46 482	11 282
8	327 384	40 923	32 925		32 925	7 998
9	274 482	30 498	24 549		24 549	5 949
10 ou mais	2 491 076	164 296	131 858		131 858	32 438
Total	14 926 599	8 500 146	7 051 481	869 312	7 920 793	579 353

- Identificaram-se **8,5 milhões de indivíduos únicos (col 3)** nos Censos 2011 onde foi possível colocar um NIC proveniente da BDIC 2010.
- Com freguesia de residência igual na BDIC e nos censos, obtiveram-se **7 milhões** (col 4) de registos – correspondem, em cada um dos ficheiros, ao mesmo indivíduo.

Nestes 7 milhões de registos existiam 2 tipos de situações:

- Freguesias iguais com registos 1 para 1 entre BDIC e os Censos – 5,4 milhões de registos (col 4, 1º valor).
- Freguesias iguais com registos 1 dos Censos para n da BDIC – 1,6 milhões de registos (col 4, soma do 2º ao 10º valor) com pelo menos um par cuja freguesia de residência é igual.
- Com freguesia de residência diferente na BDIC e nos censos, obtiveram-se **870 mil** (col 5) registos 1 para 1.
- Com freguesia de residência diferente na BDIC e nos censos, obtiveram-se **580 mil** registos (col 7) 1 para n.

Até este ponto, a freguesia de residência funcionou como variável de decisão, no entanto, para que mais registos dos censos pudessem ter um NIC foi necessário utilizar outras variáveis comuns aos dois ficheiros - as únicas variáveis passíveis de utilizar também como variáveis de decisão são o estado civil e a freguesia de naturalidade.

Assim, há necessidade de separar duas situações que se repetem sempre que se introduz novas variáveis:

1ª- Quando existiu uma decisão, isto é, para um registo dos censos existe um registo na BDIC correspondente ao mesmo indivíduo e o resultado é a afetação do NIC.

2ª- Quando não existiu decisão, ou seja, para um registo censitário não se conseguiu determinar o seu correspondente na BDIC e consequentemente o registo censitário fica sem NIC.

Em números, o processo pode ser traduzido da seguinte forma:

- Ficheiro com cerca de 7 milhões de registos – são os registos cujo matching foi de 1 para 1 e a freguesia é igual nas duas fontes somado com os registos que embora o matching seja de 1 para n há pelo menos um par onde a freguesia de residência é igual, sendo esse o escolhido.
- Ficheiro correspondente a 1,5 milhões de registos que por a freguesia de residência ser diferente, não existiu decisão. Nestes casos juntou-se uma nova variável de decisão, o estado civil.

Dessa junção resultou:

- Ficheiro com cerca de 1 milhão de registos – são os registos cujo matching foi de 1 para 1 e a freguesia é diferente nas duas fontes mas o estado civil é igual somado com os registos que embora o matching seja de 1 para n há pelo menos um par onde o estado civil é igual, sendo esse o escolhido.
- Ficheiro correspondente a 460 mil registos que por o estado civil ser diferente, não existiu decisão. Nestes casos juntou-se uma nova variável de decisão, a freguesia de naturalidade.

Dessa junção resultou:

- Ficheiro com cerca de 19 mil registos – são os registos cujo matching foi de 1 para 1 e a freguesia é diferente nas duas fontes, assim como o estado civil, mas a naturalidade é igual somado com os registos que embora o matching seja de 1 para n há pelo menos um par onde a naturalidade é igual, sendo esse o escolhido.
- Ficheiro correspondente a 440 mil registos que por a naturalidade ser diferente, não existiu decisão.

De referir que a freguesia de naturalidade nos censos corresponde à freguesia de residência da mãe à data do nascimento e portanto esta comparação pode induzir algum erro.

Em conclusão, o ficheiro onde a cada individuo único dos censos se atribuiu um NIC da BDIC com determinado nível de certeza possui cerca de 8,2 milhões de

registos. Embora todos estes registos tenham resultado de um processo de decisão é necessário obter total certeza sobre o matching dos “pares” constituídos.

Efetivamente, pode acontecer num matching que a mesma chave com a mesma freguesia de residência (ou outra variável de decisão) corresponda a indivíduos distintos, por isso é importante que mesmo nessas situações se liguem outras variáveis comuns para que o grau de certeza da decisão seja maior.

Para os 3 ficheiros (Censos 2011, BDIC 2010 e o ficheiro que resultou do matching dos 2 anteriores), analisou-se a distribuição das variáveis estado civil e idade por escalão. Verificou-se que o peso em cada grupo dos registos que efetuaram matching estão muito próximas dos outros dois ficheiros, como se pode observar nos gráficos seguintes. Estes resultados indicam uma boa opção relativamente à chave de ligação.

Fig 1 - Distribuição dos registos da BDIC 2010, Censos 2011 e Ficheiro matching por estado civil legal

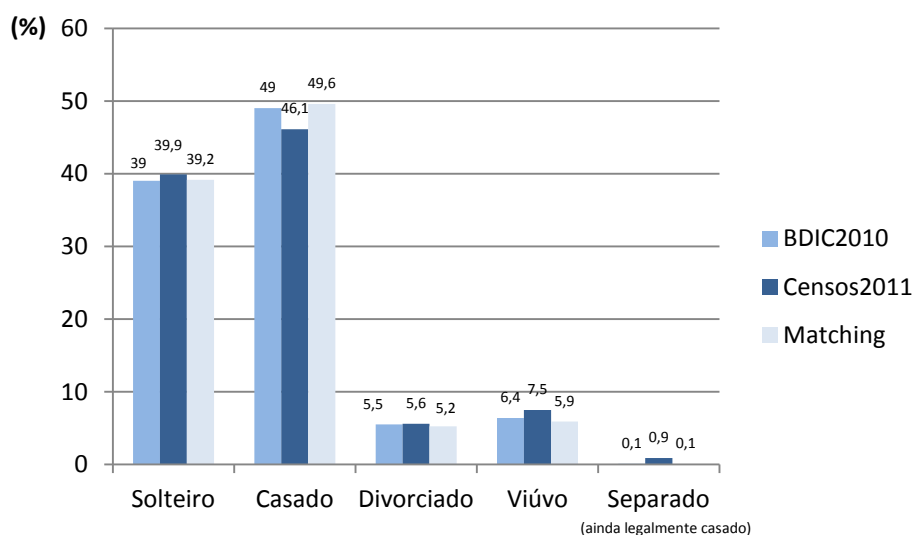
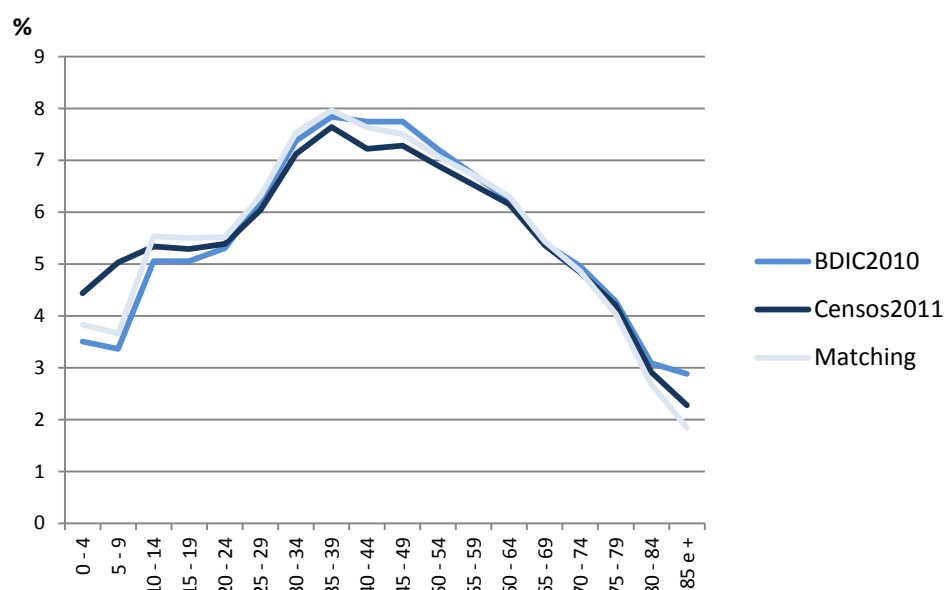


Fig 2 - Distribuição dos registos da BDIC 2010, Censos 2011 e Ficheiro matching por escalões etários



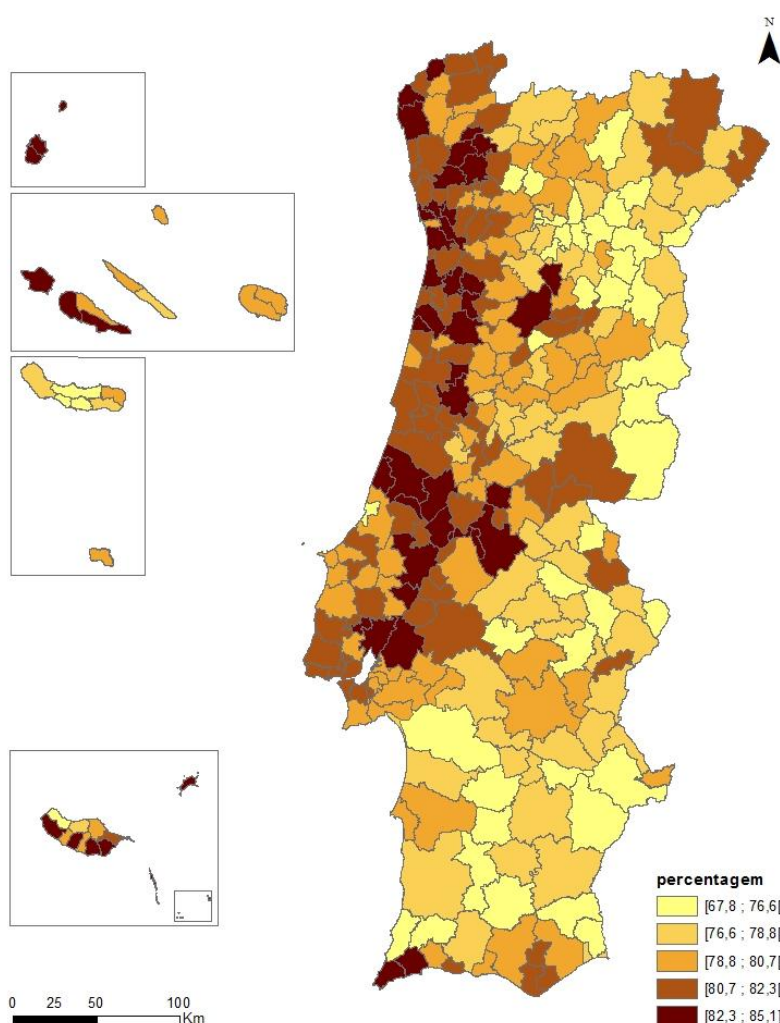
Antes de 2007 as crianças até aos 6 anos não eram incluídas na BDIC se não possuísem BI, apenas a partir de 2008 o projeto “Nascer Cidadão” permite cobrir esta faixa etária; assim, para os indivíduos dos censos que possuem entre 3 e 6 anos há uma forte possibilidade de não efetuarem matching com a BDIC 2010 - se não possuírem BI não estão incluídos na BDIC. De referir que 366 139 registos com idade inferior a 6 anos fizeram matching entre os censos 2011 e a BDIC 2010 (crianças portadoras de BI).

Relativamente ao grupo com 85 e mais anos, o facto da BDIC não abater um registo sem confirmação/comunicação do óbito pode justificar as diferenças.

No anexo 2 pode ser consultado, por município, a cobertura do ficheiro do matching face ao total da população de nacionalidade portuguesa residente em Portugal (Censos 2011) e face ao total da população de nacionalidade portuguesa com morada de residência em Portugal (BDIC 2010). Para Portugal, esses valores são de 81% e de 73%, respetivamente.

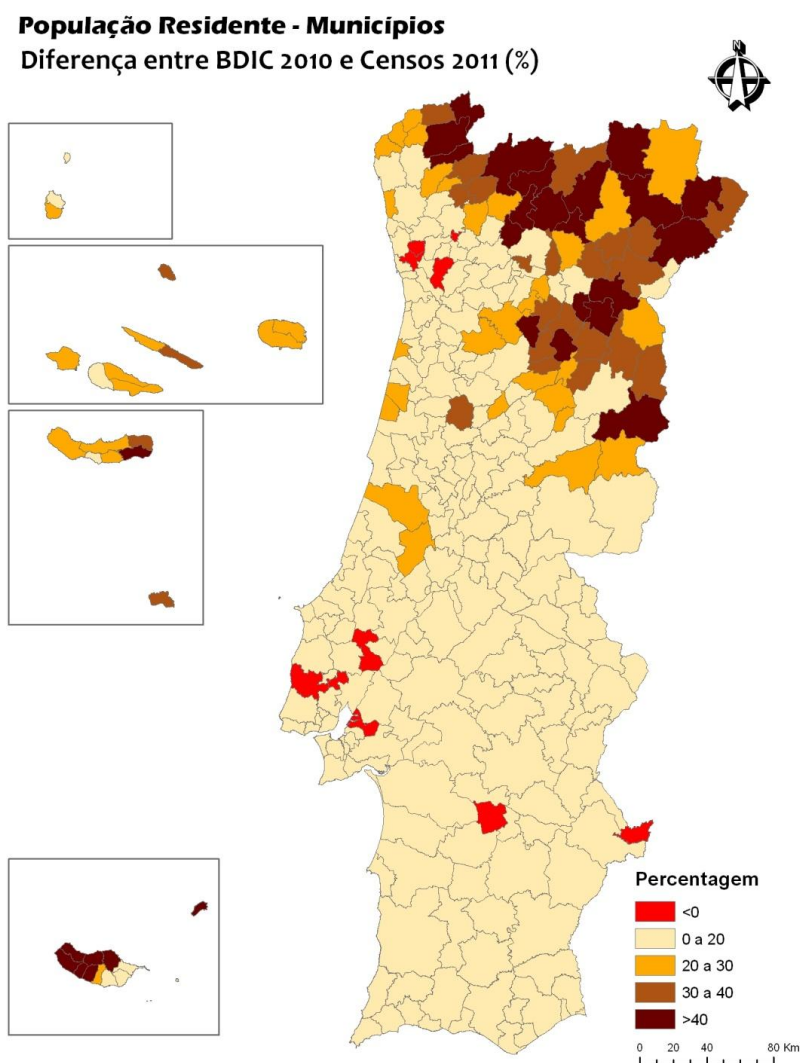
Para uma melhor visualização do principal resultado obtido (colocação de NIC nos registos censitários e cobertura desse resultado face ao ficheiro dos Censos 2011), elaborou-se um cartograma por município.

Fig 3 – Cobertura do ficheiro resultante do matching (registos censitários com NIC) face ao ficheiro dos censos 2011, por município (%)



No relatório *Análise dos ficheiros administrativos- BDIC 2010, de dezembro de 2013*, foi desenvolvido um estudo que conduziu à elaboração do cartograma seguinte.

Fig 4 – Diferença entre a BDIC 2010 e os Censos 2011, em percentagem



Da análise da figura 3, conclui-se que são os municípios localizados na faixa litoral a partir de Lisboa para Norte aqueles que apresentam uma melhor cobertura (a maioria é superior a 80%) do resultado do matching relativamente aos Censos 2011. Estes municípios situam-se, relativamente à diferença detetada entre a BDIC e os censos (figura 4) nas classes que apresentam as menores

diferenças (entre 0 e 20% ou diferenças negativas quando os censos apresentam mais população que a BDIC).

Para os municípios que apresentam as diferenças mais significativas entre os dados da BDIC e os censos, localizados na região interior norte, normalmente caracterizados por apresentarem uma população mais envelhecida, maiores deslocamentos da população para o estrangeiro ou no país, o resultado da cobertura do ficheiro de matching face aos censos varia, para a maioria, entre 68% e 79% (as duas classes inferiores).

Isto permite concluir que quando a população efetivamente reside nos municípios que “administrativamente” declara, esta chave de ligação conduz a bons resultados. No entanto, é possível melhorar as técnicas de matching, já que para uma situação de matching próximo dos 100%, o mapa obtido para a figura 3 deveria ser próximo da distribuição do mapa obtido na figura 4, sendo as regiões do interior norte aquelas que apresentariam as menores taxas de matching.

Para caracterizar a população de cada uma das fontes que não efetuou matching focamo-nos apenas no grupo de registos entre os 6 e os 100 anos. Para estes, efetuou-se um conjunto de apuramentos relacionados com:

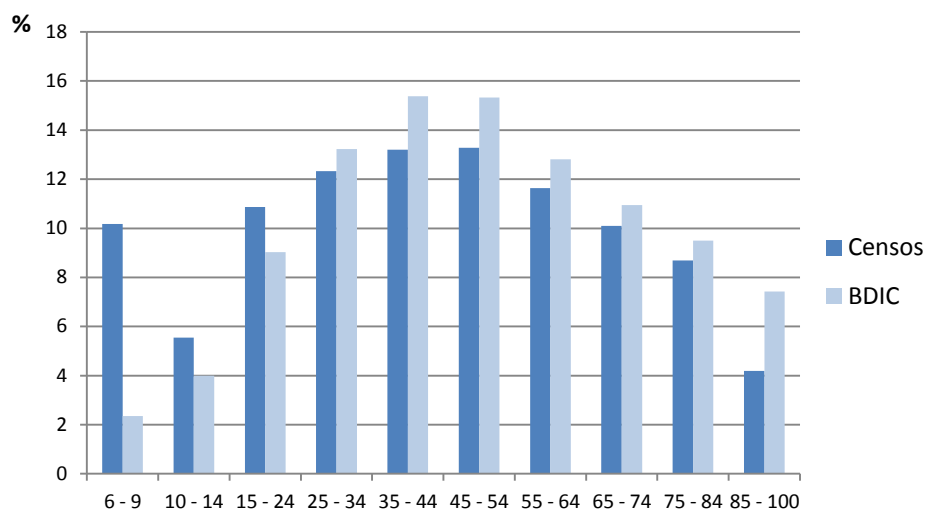
- Classes de idade
- Sexo
- Condição Perante o Trabalho, apenas para os Censos 2011
- Nível instrução (a frequentar), apenas para os Censos 2011

Os resultados obtidos, agregados para Portugal são os seguintes:

Quadro 2 - Distribuição por classes de idade dos registos que não efetuaram matching nos Censos 2011 e na BDIC 2010

	Classes de idade (anos)										
	Total	6-9	10-14	15-24	25-34	35-44	45-54	55-64	65-74	75-84	85-100
Censos	1 502 740	152 956	83 324	163 403	185 291	198 290	199 535	174 854	151 669	130 510	62 908
BDIC	2 864 784	67 531	114 662	258 639	379 066	440 436	439 117	366 845	313 680	271 954	212 854

Fig 5 - Distribuição por classes de idade dos registos que não efetuaram matching nos Censos 2011 e na BDIC 2010



Por idades, e pelos motivos já apresentados relacionados com a inclusão de cidadãos na BDIC, uma grande parte de indivíduos dos censos com 6 anos não puderam ser detetados na BDIC; a importância do grupo com idades entre os 25 e os 65 anos que “sobrou” na BDIC relaciona-se, provavelmente, com indivíduos que não residem em Portugal.

Quadro 4 - Distribuição por sexo dos registos que não efetuaram matching nos Censos 2011 e na BDIC 2010

	Sexo		
	Total	H	M
Censos	1 502 740	638 726	864 014
BDIC	2 864 784	1 449 579	1 415 205

Quadro 5 - Distribuição por condição perante o trabalho dos registos que não efetuaram matching nos Censos 2011

CPT								
Total	Empregados	Desempregados	<15 anos	Estudantes	Domésticos	Reformados	Incapacitados	Outros
1 502 740	524 079	91 878	236 280	88 223	73 490	387 098	27 053	74 639

Ao analisar os registos que não efetuaram matching nos Censos 2011, regista-se que o valor mais elevado se refere à população empregada, o que corrobora os dados obtidos anteriormente relativos às idades.

Quadro 6 - Distribuição por frequência do nível de ensino dos registos que não efetuaram matching nos Censos 2011

	A frequentar
Total	366 697
Frequentar o pré-escolar	12 092
1º ciclo	162 065
2º ciclo	37 412
3º ciclo	58 887
Secundário	51 551
Pós-secundário	2 769
Licenciatura	30 111
Mestrado	10 045
Doutoramento	1 765

Como o primeiro objetivo deste trabalho era colocar um NIC em cada registo dos censos, nos cerca de 1.5 milhões de registos que “sobraram”, efetuou-se a comparação com outras fontes:

- Segurança Social (SS)
- Educação
- Caixa Geral de Aposentações (CGA)

Para a segurança social e para a educação o matching foi efetuado através da mesma chave de ligação que foi utilizada na comparação entre os Censos 2011 e a BDIC 2010: data de nascimento/3 primeiras letras do 1º nome e 3 últimas letras do último nome/sexo. Também se aplicaram as mesmas variáveis de decisão – freguesia de residência, estado civil e naturalidade.

Para a CGA, atendendo a que não existe nome neste ficheiro, o matching foi efetuado através de uma chave de ligação constituída pela variável sexo e pela variável data de nascimento. Obteve-se um resultado de cerca de 1000 registos ligados, mas este resultado não foi considerado aceitável devido à fraca confiança na chave de ligação.

Quadro 7 - Resultado da comparação com outras fontes do ficheiro dos censos 2011 que não efetuou matching com a BDIC 2010

Registos dos Censos 2011 que não efetuaram matching c/ BDIC 2010: 1 502 740	Efetuou matching através de chave de ligação c/			
	Total	SS	Educação	Ambos
	241 747	110 310	111 739	19 698
		c/ NIC: 106 403	c/ NIC:60 217	c/NIC:18 512

Constata-se que 96% (106 403) dos registos que fizeram matching com a SS possuem NIC mas, para a Educação, o valor é de apenas 54% (60 217).

Anexo 2 - Cobertura do ficheiro do matching face ao total do ficheiro dos censos 2011 e da BDIC 2010, por município

Código	Município	Censos 2011 – população portuguesa	BDIC 2010	Ficheiro após matching	Ficheiro matching /Censos 2011	Ficheiro matching /BDIC 2010	Variação matching /Censos 2011	Variação matching /BDIC 2010
TOTAL		10167129	11311392	8242088	81,1	72,9	-18,9	-27,1
0101	Águeda	46642	51216	38456	82,4	75,1	-17,6	-24,9
0102	Albergaria-a-Velha	24713	27309	20181	81,7	73,9	-18,3	-26,1
0103	Anadia	28426	33958	22965	80,8	67,6	-19,2	-32,4
0104	Arouca	22198	25126	18146	81,7	72,2	-18,3	-27,8
0105	Aveiro	75797	81855	62909	83,0	76,9	-17,0	-23,1
0106	Castelo de Paiva	16674	17564	13381	80,3	76,2	-19,7	-23,8
0107	Espinho	31331	35601	25502	81,4	71,6	-18,6	-28,4
0108	Estarreja	26672	29946	21516	80,7	71,8	-19,3	-28,2
0109	Feira	137619	148711	112450	81,7	75,6	-18,3	-24,4
0110	Ílhavo	37609	43458	30980	82,4	71,3	-17,6	-28,7
0111	Mealhada	20051	22108	16555	82,6	74,9	-17,4	-25,1
0112	Murtosa	10466	13346	8449	80,7	63,3	-19,3	-36,7
0113	Oliveira de Azeméis	67856	72409	56657	83,5	78,2	-16,5	-21,8
0114	Oliveira do Bairro	22351	25349	17982	80,5	70,9	-19,5	-29,1
0115	Ovar	54682	59163	45432	83,1	76,8	-16,9	-23,2
0116	São João da Madeira	21212	23299	17741	83,6	76,1	-16,4	-23,9
0117	Sever do Vouga	12170	14054	10032	82,4	71,4	-17,6	-28,6
0118	Vagos	22350	27767	18183	81,4	65,5	-18,6	-34,5
0119	Vale de Cambra	22546	26518	19178	85,1	72,3	-14,9	-27,7
0201	Aljustrel	9156	10346	6976	76,2	67,4	-23,8	-32,6
0202	Almodôvar	7322	8114	5531	75,5	68,2	-24,5	-31,8
0203	Alvito	2405	2407	1856	77,2	77,1	-22,8	-22,9
0204	Barrancos	1803	1753	1431	79,4	81,6	-20,6	-18,4
0205	Beja	34888	36802	27460	78,7	74,6	-21,3	-25,4
0206	Castro Verde	7128	7881	5457	76,6	69,2	-23,4	-30,8
0207	Cuba	4699	4833	3618	77,0	74,9	-23,0	-25,1
0208	Ferreira do Alentejo	7969	8561	6097	76,5	71,2	-23,5	-28,8
0209	Mértola	7150	8111	5574	78,0	68,7	-22,0	-31,3
0210	Moura	14919	16769	11163	74,8	66,6	-25,2	-33,4
0211	Odemira	23662	24845	18145	76,7	73,0	-23,3	-27,0
0212	Ourique	5179	5652	3928	75,8	69,5	-24,2	-30,5
0213	Serpa	15324	16699	11642	76,0	69,7	-24,0	-30,3
0214	Vidigueira	5717	6211	4198	73,4	67,6	-26,6	-32,4
0301	Amares	18696	23803	15428	82,5	64,8	-17,5	-35,2
0302	Barcelos	119427	127788	97152	81,3	76,0	-18,7	-24,0
0303	Braga	176691	188617	147276	83,4	78,1	-16,6	-21,9
0304	Cabeceiras de Basto	16579	20495	13173	79,5	64,3	-20,5	-35,7
0305	Celorico de Basto	19980	23098	15036	75,3	65,1	-24,7	-34,9
0306	Esposende	33824	40857	27582	81,5	67,5	-18,5	-32,5
0307	Fafe	50248	61066	41039	81,7	67,2	-18,3	-32,8
0308	Guimarães	156577	170593	129669	82,8	76,0	-17,2	-24,0
0309	Póvoa de Lanhoso	21652	29370	17865	82,5	60,8	-17,5	-39,2
0310	Terras de Bouro	7199	9426	5827	80,9	61,8	-19,1	-38,2

0311	Vieira do Minho	12852	17525	9917	77,2	56,6	-22,8	-43,4
0312	Vila Nova de Famalicão	132304	139117	109889	83,1	79,0	-16,9	-21,0
0313	Vila Verde	47387	57781	37849	79,9	65,5	-20,1	-34,5
0314	Vizela	23574	23384	19343	82,1	82,7	-17,9	-17,3
0401	Alfândega da Fé	5037	6997	3902	77,5	55,8	-22,5	-44,2
0402	Bragança	34755	44730	28526	82,1	63,8	-17,9	-36,2
0403	Carrazeda de Ansiães	6233	8660	4654	74,7	53,7	-25,3	-46,3
0404	Freixo de Espada à Cinta	3689	4458	2500	67,8	56,1	-32,2	-43,9
0405	Macedo de Cavaleiros	15586	22920	12591	80,8	54,9	-19,2	-45,1
0406	Miranda do Douro	7359	9699	5982	81,3	61,7	-18,7	-38,3
0407	Mirandela	23405	29219	18391	78,6	62,9	-21,4	-37,1
0408	Mogadouro	9424	13591	7337	77,9	54,0	-22,1	-46,0
0409	Torre de Moncorvo	8498	12046	6583	77,5	54,6	-22,5	-45,4
0410	Vila Flor	6540	9168	4900	74,9	53,4	-25,1	-46,6
0411	Vimioso	4617	7780	3620	78,4	46,5	-21,6	-53,5
0412	Vinhais	8996	13837	6898	76,7	49,9	-23,3	-50,1
0501	Belmonte	6787	8113	5267	77,6	64,9	-22,4	-35,1
0502	Castelo Branco	55079	61284	44476	80,7	72,6	-19,3	-27,4
0503	Covilhã	51204	59603	40918	79,9	68,7	-20,1	-31,3
0504	Fundão	28742	35514	22446	78,1	63,2	-21,9	-36,8
0505	Idanha-a-Nova	9524	11571	7210	75,7	62,3	-24,3	-37,7
0506	Oleiros	5655	6432	4340	76,7	67,5	-23,3	-32,5
0507	Penamacor	5633	7139	4231	75,1	59,3	-24,9	-40,7
0508	Proença-a-Nova	8213	9793	6669	81,2	68,1	-18,8	-31,9
0509	Sertão	15647	17547	12532	80,1	71,4	-19,9	-28,6
0510	Vila de Rei	3412	3543	2813	82,4	79,4	-17,6	-20,6
0511	Vila Velha de Ródão	3472	3634	2805	80,8	77,2	-19,2	-22,8
0601	Arganil	11787	13079	9358	79,4	71,5	-20,6	-28,5
0602	Cantanhede	36023	42805	29089	80,8	68,0	-19,2	-32,0
0603	Coimbra	139999	147562	115139	82,2	78,0	-17,8	-22,0
0604	Condeixa-a-Nova	16766	16917	13608	81,2	80,4	-18,8	-19,6
0605	Figueira da Foz	60704	68979	49637	81,8	72,0	-18,2	-28,0
0606	Góis	4159	4392	3209	77,2	73,1	-22,8	-26,9
0607	Lousã	17096	18174	13638	79,8	75,0	-20,2	-25,0
0608	Mira	12255	15480	9735	79,4	62,9	-20,6	-37,1
0609	Miranda do Corvo	12869	13310	10382	80,7	78,0	-19,3	-22,0
0610	Montemor-o-Velho	25898	26429	21166	81,7	80,1	-18,3	-19,9
0611	Oliveira do Hospital	20445	22351	16063	78,6	71,9	-21,4	-28,1
0612	Pampilhosa da Serra	4451	4895	3427	77,0	70,0	-23,0	-30,0
0613	Penacova	15116	17452	12176	80,6	69,8	-19,4	-30,2
0614	Penela	5830	6199	4650	79,8	75,0	-20,2	-25,0
0615	Soure	19009	21226	15612	82,1	73,6	-17,9	-26,4
0616	Tábua	11646	12824	9225	79,2	71,9	-20,8	-28,1
0617	Vila Nova de Poiares	7057	7717	5541	78,5	71,8	-21,5	-28,2
0701	Alandroal	5757	6072	4482	77,9	73,8	-22,1	-26,2
0702	Arraiolos	7248	7433	5816	80,2	78,2	-19,8	-21,8
0703	Borba	7200	7516	5619	78,0	74,8	-22,0	-25,2
0704	Estremoz	14048	14668	10485	74,6	71,5	-25,4	-28,5
0705	Évora	55103	56969	43832	79,5	76,9	-20,5	-23,1
0706	Montemor-o-Novo	17054	18089	13198	77,4	73,0	-22,6	-27,0
0707	Mora	4915	5432	3751	76,3	69,1	-23,7	-30,9
0708	Mourão	2629	2957	2022	76,9	68,4	-23,1	-31,6
0709	Portel	6345	6617	4937	77,8	74,6	-22,2	-25,4

0710	Redondo	6876	7107	5448	79,2	76,7	-20,8	-23,3
0711	Reguengos de Monsaraz	10555	10959	8292	78,6	75,7	-21,4	-24,3
0712	Vendas Novas	11612	12493	9229	79,5	73,9	-20,5	-26,1
0713	Viana do Alentejo	5620	5941	4356	77,5	73,3	-22,5	-26,7
0714	Vila Viçosa	8218	8607	6655	81,0	77,3	-19,0	-22,7
0801	Albufeira	33656	37039	27103	80,5	73,2	-19,5	-26,8
0802	Alcoutim	2863	3347	2187	76,4	65,3	-23,6	-34,7
0803	Aljezur	4610	4877	3462	75,1	71,0	-24,9	-29,0
0804	Castro Marim	6276	7004	4740	75,5	67,7	-24,5	-32,3
0805	Faro	59236	66751	48658	82,1	72,9	-17,9	-27,1
0806	Lagoa	20390	21478	16475	80,8	76,7	-19,2	-23,3
0807	Lagos	25806	27427	21693	84,1	79,1	-15,9	-20,9
0808	Loulé	60297	67028	48137	79,8	71,8	-20,2	-28,2
0809	Monchique	5523	5999	4186	75,8	69,8	-24,2	-30,2
0810	Olhão	42483	44886	34419	81,0	76,7	-19,0	-23,3
0811	Portimão	49163	53237	39585	80,5	74,4	-19,5	-25,6
0812	São Brás de Alportel	9577	10890	7796	81,4	71,6	-18,6	-28,4
0813	Silves	32925	35295	25349	77,0	71,8	-23,0	-28,2
0814	Tavira	23693	25927	19013	80,2	73,3	-19,8	-26,7
0815	Vila do Bispo	4457	4711	3684	82,7	78,2	-17,3	-21,8
0816	Vila Real de Santo António	17861	20276	13876	77,7	68,4	-22,3	-31,6
0901	Aguiar da Beira	5408	7926	4099	75,8	51,7	-24,2	-48,3
0902	Almeida	7096	9828	5571	78,5	56,7	-21,5	-43,3
0903	Celorico da Beira	7624	10751	5940	77,9	55,3	-22,1	-44,7
0904	Figueira de Castelo Rodrigo	6161	7773	4739	76,9	61,0	-23,1	-39,0
0905	Fornos de Algodres	4954	6413	4017	81,1	62,6	-18,9	-37,4
0906	Gouveia	13865	18290	11140	80,3	60,9	-19,7	-39,1
0907	Guarda	41973	49218	33413	79,6	67,9	-20,4	-32,1
0908	Manteigas	3408	4310	2722	79,9	63,2	-20,1	-36,8
0909	Meda	5140	7402	3901	75,9	52,7	-24,1	-47,3
0910	Pinhel	9532	12622	7149	75,0	56,6	-25,0	-43,4
0911	Sabugal	12390	19957	9449	76,3	47,3	-23,7	-52,7
0912	Seia	24415	29339	19587	80,2	66,8	-19,8	-33,2
0913	Trancoso	9746	12880	7053	72,4	54,8	-27,6	-45,2
0914	Vila Nova de Foz Côa	7145	10288	5427	76,0	52,8	-24,0	-47,2
1001	Alcobaça	55188	57945	44425	80,5	76,7	-19,5	-23,3
1002	Alvaiázere	7121	8466	5713	80,2	67,5	-19,8	-32,5
1003	Ansião	12893	14679	10055	78,0	68,5	-22,0	-31,5
1004	Batalha	15264	17187	12727	83,4	74,1	-16,6	-25,9
1005	Bombarral	12739	14676	10281	80,7	70,1	-19,3	-29,9
1006	Caldas da Rainha	49598	53638	40231	81,1	75,0	-18,9	-25,0
1007	Castanheira de Pêra	3129	3476	2454	78,4	70,6	-21,6	-29,4
1008	Figueiró dos Vinhos	6084	7173	4934	81,1	68,8	-18,9	-31,2
1009	Leiria	122482	134991	102409	83,6	75,9	-16,4	-24,1
1010	Marinha Grande	37519	39295	31288	83,4	79,6	-16,6	-20,4
1011	Nazaré	14691	17102	11187	76,1	65,4	-23,9	-34,6
1012	Óbidos	11449	12395	9197	80,3	74,2	-19,7	-25,8
1013	Pedrogão Grande	3762	4262	3059	81,3	71,8	-18,7	-28,2
1014	Peniche	26950	30647	21678	80,4	70,7	-19,6	-29,3
1015	Pombal	53845	66581	43976	81,7	66,0	-18,3	-34,0
1016	Porto de Mós	23646	26075	19340	81,8	74,2	-18,2	-25,8
1101	Alenquer	40492	41193	32849	81,1	79,7	-18,9	-20,3
1102	Arruda dos Vinhos	12701	12224	10598	83,4	86,7	-16,6	-13,3

1103	Azambuja	20643	20675	16436	79,6	79,5	-20,4	-20,5
1104	Cadaval	13895	15261	11185	80,5	73,3	-19,5	-26,7
1105	Cascais	187794	205622	151960	80,9	73,9	-19,1	-26,1
1106	Lisboa	513197	604983	415444	81,0	68,7	-19,0	-31,3
1107	Loures	187404	205697	149735	79,9	72,8	-20,1	-27,2
1108	Lourinhã	24677	27053	19664	79,7	72,7	-20,3	-27,3
1109	Mafra	72698	70848	58950	81,1	83,2	-18,9	-16,8
1110	Oeiras	161920	174357	132090	81,6	75,8	-18,4	-24,2
1111	Sintra	342808	364445	277859	81,1	76,2	-18,9	-23,8
1112	Sobral de Monte Agraço	9701	9764	7849	80,9	80,4	-19,1	-19,6
1113	Torres Vedras	76239	78981	61338	80,5	77,7	-19,5	-22,3
1114	Vila Franca de Xira	129128	130838	106686	82,6	81,5	-17,4	-18,5
1115	Amadora	156236	166573	127219	81,4	76,4	-18,6	-23,6
1116	Odivelas	131598	132202	107931	82,0	81,6	-18,0	-18,4
1201	Alter do Chão	3444	3641	2622	76,1	72,0	-23,9	-28,0
1202	Arronches	3095	3277	2387	77,1	72,8	-22,9	-27,2
1203	Avis	4402	4510	3435	78,0	76,2	-22,0	-23,8
1204	Campo Maior	8191	9216	6190	75,6	67,2	-24,4	-32,8
1205	Castelo de Vide	3315	3513	2528	76,3	72,0	-23,7	-28,0
1206	Crato	3640	3977	2799	76,9	70,4	-23,1	-29,6
1207	Elvas	22326	24647	17277	77,4	70,1	-22,6	-29,9
1208	Fronteira	3335	3487	2616	78,4	75,0	-21,6	-25,0
1209	Gavião	4106	4400	3220	78,4	73,2	-21,6	-26,8
1210	Marvão	3384	3658	2684	79,3	73,4	-20,7	-26,6
1211	Monforte	3288	3410	2467	75,0	72,3	-25,0	-27,7
1212	Nisa	7351	8121	5764	78,4	71,0	-21,6	-29,0
1213	Ponte de Sor	16426	17580	12913	78,6	73,5	-21,4	-26,5
1214	Portalegre	24432	25634	19963	81,7	77,9	-18,3	-22,1
1215	Sousel	4991	5191	3845	77,0	74,1	-23,0	-25,9
1301	Amarante	55775	64981	44827	80,4	69,0	-19,6	-31,0
1302	Baião	20459	21919	16494	80,6	75,2	-19,4	-24,8
1303	Felgueiras	57641	61452	46507	80,7	75,7	-19,3	-24,3
1304	Gondomar	166419	167711	139024	83,5	82,9	-16,5	-17,1
1305	Lousada	47126	47946	37874	80,4	79,0	-19,6	-21,0
1306	Maia	133046	132395	111426	83,7	84,2	-16,3	-15,8
1307	Marco de Canaveses	53061	56537	43070	81,2	76,2	-18,8	-23,8
1308	Matosinhos	172220	174310	142651	82,8	81,8	-17,2	-18,2
1309	Paços de Ferreira	55921	55976	45037	80,5	80,5	-19,5	-19,5
1310	Paredes	86376	86076	70066	81,1	81,4	-18,9	-18,6
1311	Penafiel	71866	73786	58840	81,9	79,7	-18,1	-20,3
1312	Porto	230596	250308	185665	80,5	74,2	-19,5	-25,8
1313	Póvoa de Varzim	62321	71643	51024	81,9	71,2	-18,1	-28,8
1314	Santo Tirso	70908	76657	58098	81,9	75,8	-18,1	-24,2
1315	Valongo	92822	94768	76761	82,7	81,0	-17,3	-19,0
1316	Vila do Conde	78267	81285	64021	81,8	78,8	-18,2	-21,2
1317	Vila Nova de Gaia	297203	301495	245957	82,8	81,6	-17,2	-18,4
1318	Trofa	38347	37185	31300	81,6	84,2	-18,4	-15,8
1401	Abrantes	38801	41518	32014	82,5	77,1	-17,5	-22,9
1402	Alcanena	13480	15216	11075	82,2	72,8	-17,8	-27,2
1403	Almeirim	22669	24146	18343	80,9	76,0	-19,1	-24,0
1404	Alpiarça	7503	7867	6114	81,5	77,7	-18,5	-22,3
1405	Benavente	26959	27568	22417	83,2	81,3	-16,8	-18,7
1406	Cartaxo	23611	24893	19438	82,3	78,1	-17,7	-21,9

1407	Chamusca	10024	10400	8067	80,5	77,6	-19,5	-22,4
1408	Constância	4018	4079	3356	83,5	82,3	-16,5	-17,7
1409	Coruche	19595	21090	15826	80,8	75,0	-19,2	-25,0
1410	Entroncamento	19711	20148	16342	82,9	81,1	-17,1	-18,9
1411	Ferreira do Zêzere	8407	9531	6661	79,2	69,9	-20,8	-30,1
1412	Golegã	5410	5481	4400	81,3	80,3	-18,7	-19,7
1413	Mação	7253	8191	5838	80,5	71,3	-19,5	-28,7
1414	Rio Maior	20457	21566	16272	79,5	75,5	-20,5	-24,5
1415	Salvaterra de Magos	21668	22536	17480	80,7	77,6	-19,3	-22,4
1416	Santarém	60148	64578	49807	82,8	77,1	-17,2	-22,9
1417	Sardoal	3892	4030	3160	81,2	78,4	-18,8	-21,6
1418	Tomar	39883	44111	32304	81,0	73,2	-19,0	-26,8
1419	Torres Novas	36170	38611	30030	83,0	77,8	-17,0	-22,2
1420	Vila Nova da Barquinha	7247	7610	6023	83,1	79,1	-16,9	-20,9
1421	Ourém	44323	54169	37093	83,7	68,5	-16,3	-31,5
1501	Alcácer do Sal	12726	13139	9538	74,9	72,6	-25,1	-27,4
1502	Alcochete	16782	16626	13519	80,6	81,3	-19,4	-18,7
1503	Almada	162620	174842	131556	80,9	75,2	-19,1	-24,8
1504	Barreiro	75326	81890	61226	81,3	74,8	-18,7	-25,2
1505	Grândola	14102	14504	11001	78,0	75,8	-22,0	-24,2
1506	Moita	62744	70392	50062	79,8	71,1	-20,2	-28,9
1507	Montijo	47927	48511	38354	80,0	79,1	-20,0	-20,9
1508	Palmela	60621	61310	48441	79,9	79,0	-20,1	-21,0
1509	Santiago do Cacém	28616	29684	22753	79,5	76,7	-20,5	-23,3
1510	Seixal	147605	156572	120129	81,4	76,7	-18,6	-23,3
1511	Sesimbra	47269	48054	37475	79,3	78,0	-20,7	-22,0
1512	Setúbal	114396	123725	92181	80,6	74,5	-19,4	-25,5
1513	Sines	13345	13814	10750	80,6	77,8	-19,4	-22,2
1601	Arcos de Valdevez	22542	35029	18177	80,6	51,9	-19,4	-48,1
1602	Caminha	16289	20055	13399	82,3	66,8	-17,7	-33,2
1603	Melgaço	8967	14526	7367	82,2	50,7	-17,8	-49,3
1604	Monção	18701	25696	15319	81,9	59,6	-18,1	-40,4
1605	Paredes de Coura	9066	11597	7211	79,5	62,2	-20,5	-37,8
1606	Ponte da Barca	11889	17549	9509	80,0	54,2	-20,0	-45,8
1607	Ponte de Lima	43091	51552	34661	80,4	67,2	-19,6	-32,8
1608	Valença	13605	16768	11250	82,7	67,1	-17,3	-32,9
1609	Viana do Castelo	87585	104483	72569	82,9	69,5	-17,1	-30,5
1610	Vila Nova de Cerveira	8975	10904	7323	81,6	67,2	-18,4	-32,8
1701	Alijó	11841	15433	9073	76,6	58,8	-23,4	-41,2
1702	Boticas	5711	9324	4489	78,6	48,1	-21,4	-51,9
1703	Chaves	40540	55177	32263	79,6	58,5	-20,4	-41,5
1704	Mesão Frio	4401	5222	3358	76,3	64,3	-23,7	-35,7
1705	Mondim de Basto	7456	10598	5586	74,9	52,7	-25,1	-47,3
1706	Montalegre	10428	18261	8012	76,8	43,9	-23,2	-56,1
1707	Murça	5916	8589	4725	79,9	55,0	-20,1	-45,0
1708	Peso da Régua	17006	20378	13004	76,5	63,8	-23,5	-36,2
1709	Ribeira de Pena	6505	10446	5226	80,3	50,0	-19,7	-50,0
1710	Sabrosa	6296	8400	4765	75,7	56,7	-24,3	-43,3
1711	Santa Marta de Penaguião	7235	9859	5617	77,6	57,0	-22,4	-43,0
1712	Valpaços	16659	25304	12539	75,3	49,6	-24,7	-50,4
1713	Vila Pouca de Aguiar	13047	21068	10493	80,4	49,8	-19,6	-50,2
1714	Vila Real	51218	61250	40905	79,9	66,8	-20,1	-33,2
1801	Armamar	6252	7738	4648	74,3	60,1	-25,7	-39,9

1802	Carregal do Sal	9683	12323	7842	81,0	63,6	-19,0	-36,4
1803	Castro Daire	15183	19798	11709	77,1	59,1	-22,9	-40,9
1804	Cinfães	20344	21904	16133	79,3	73,7	-20,7	-26,3
1805	Lamego	26537	30374	20865	78,6	68,7	-21,4	-31,3
1806	Mangualde	19474	24546	15924	81,8	64,9	-18,2	-35,1
1807	Moimenta da Beira	10047	14147	7877	78,4	55,7	-21,6	-44,3
1808	Mortágua	9386	12325	7505	80,0	60,9	-20,0	-39,1
1809	Nelas	13858	16678	10580	76,3	63,4	-23,7	-36,6
1810	Oliveira de Frades	10062	11139	8151	81,0	73,2	-19,0	-26,8
1811	Penalva do Castelo	7903	10652	6389	80,8	60,0	-19,2	-40,0
1812	Penedono	2933	4748	2351	80,2	49,5	-19,8	-50,5
1813	Resende	11307	13424	8901	78,7	66,3	-21,3	-33,7
1814	Santa Comba Dão	11423	13791	9150	80,1	66,3	-19,9	-33,7
1815	São João da Pesqueira	7743	9397	5899	76,2	62,8	-23,8	-37,2
1816	São Pedro do Sul	16666	20697	13413	80,5	64,8	-19,5	-35,2
1817	Sátão	12306	17114	9808	79,7	57,3	-20,3	-42,7
1818	Sernancelhe	5625	7879	4382	77,9	55,6	-22,1	-44,4
1819	Tabuaço	6292	7391	4797	76,2	64,9	-23,8	-35,1
1820	Tarouca	7980	9826	5990	75,1	61,0	-24,9	-39,0
1821	Tondela	28686	34530	22797	79,5	66,0	-20,5	-34,0
1822	Vila Nova de Paiva	5112	8633	4258	83,3	49,3	-16,7	-50,7
1823	Viseu	97263	117462	80802	83,1	68,8	-16,9	-31,2
1824	Vouzela	10473	11894	8267	78,9	69,5	-21,1	-30,5
3101	Calheta (R.A.M.)	11166	16657	9300	83,3	55,8	-16,7	-44,2
3102	Câmara de Lobos	35310	42908	28154	79,7	65,6	-20,3	-34,4
3103	Funchal	109104	131661	90487	82,9	68,7	-17,1	-31,3
3104	Machico	21589	25856	17639	81,7	68,2	-18,3	-31,8
3105	Ponta do Sol	8671	12813	6904	79,6	53,9	-20,4	-46,1
3106	Porto Moniz	2694	4383	2039	75,7	46,5	-24,3	-53,5
3107	Ribeira Brava	13163	18679	10846	82,4	58,1	-17,6	-41,9
3108	Santa Cruz	41966	47442	35555	84,7	74,9	-15,3	-25,1
3109	Santana	7626	10789	6053	79,4	56,1	-20,6	-43,9
3110	São Vicente	5635	8344	4387	77,9	52,6	-22,1	-47,4
3201	Porto Santo	5231	7695	4340	83,0	56,4	-17,0	-43,6
4101	Vila do Porto	5461	7529	4378	80,2	58,1	-19,8	-41,9
4201	Lagoa	14300	16870	10730	75,0	63,6	-25,0	-36,4
4202	Nordeste	4893	6726	3865	79,0	57,5	-21,0	-42,5
4203	Ponta Delgada	68121	85825	52610	77,2	61,3	-22,8	-38,7
4204	Povoação	6288	9578	4876	77,5	50,9	-22,5	-49,1
4205	Ribeira Grande	31885	38776	24227	76,0	62,5	-24,0	-37,5
4206	Vila Franca do Campo	11126	14123	8429	75,8	59,7	-24,2	-40,3
4301	Angra do Heroísmo	34876	43623	27768	79,6	63,7	-20,4	-36,3
4302	Vila da Praia da Vitória	20816	25596	16702	80,2	65,3	-19,8	-34,7
4401	Santa Cruz da Graciosa (R.A.A.)	4310	5988	3456	80,2	57,7	-19,8	-42,3
4501	Calheta (R.A.A.)	3733	5103	2906	77,8	56,9	-22,2	-43,1
4502	Velas (R.A.A.)	5299	6793	4183	78,9	61,6	-21,1	-38,4
4601	Lajes do Pico	4587	5751	3791	82,6	65,9	-17,4	-34,1
4602	Madalena	5868	7084	4936	84,1	69,7	-15,9	-30,3
4603	São Roque do Pico	3266	4249	2607	79,8	61,4	-20,2	-38,6
4701	Horta	14577	17698	12376	84,9	69,9	-15,1	-30,1
4801	Lajes das Flores	1431	1795	1183	82,7	65,9	-17,3	-34,1
4802	Santa Cruz das Flores	2194	2614	1855	84,5	71,0	-15,5	-29,0
4901	Corvo	391	409	330	84,4	80,7	-15,6	-19,3

Anexo 3 - Alguns conceitos e disposições, incluídos no Código do Imposto Municipal sobre Imóveis (CIMI)

Prédios urbanos (Artigo 4.º): São todos aqueles que não devam ser classificados como rústicos, sem prejuízo do disposto no artigo seguinte (Nota: referente aos prédios mistos, de interesse marginal neste contexto).

Assim, importa verificar o conceito de prédio rústico: São prédios rústicos os terrenos situados fora de um aglomerado urbano que não sejam de classificar como terrenos para construção, nos termos do n.º 3 do artigo 6.º, desde que: a) Estejam afetos ou, na falta de concreta afetação, tenham como destino normal uma utilização geradora de rendimentos agrícolas, tais como são considerados para efeitos do imposto sobre o rendimento das pessoas singulares (IRS); b) Não tendo a afetação indicada na alínea anterior, não se encontrem construídos ou disponham apenas de edifícios ou construções de carácter acessório, sem autonomia económica e de reduzido valor. São ainda prédios rústicos os edifícios e construções diretamente afetos à produção de rendimentos agropecuários, quando situados nos terrenos referidos nos números anteriores.

Os **prédios urbanos** dividem-se em: a) Habitacionais; b) Comerciais, industriais ou para serviços;

c) Terrenos para construção; d) Outros.

Citam-se (partes de) outros artigos do CIMI, relevantes para efeitos deste relatório.

Conceito de matrizes prediais (Artigo 12º): **1** - As matrizes prediais são registos de que constam, designadamente, a caracterização dos prédios, a localização e o seu valor patrimonial tributário, a identidade dos proprietários e, sendo caso disso, dos usufrutuários e superficiários.

3 - Cada andar ou parte de prédio suscetível de utilização independente é considerado separadamente na inscrição matricial, a qual discrimina também o

respetivo valor patrimonial tributário. **4** - As matrizes são atualizadas anualmente com referência a 31 de Dezembro.

Inscrição nas matrizes (Artigo 13º): **1** - A inscrição de prédios na matriz e a atualização desta são efetuadas com base em declaração apresentada pelo sujeito passivo, no prazo de 60 dias contados a partir da ocorrência de qualquer dos seguintes factos: a) Uma dada realidade física passar a ser considerada como prédio; b) Verificar-se um evento suscetível de determinar uma alteração da classificação de um prédio; c) Modificarem-se os limites de um prédio; d) Concluírem-se obras de edificação, de melhoramento ou outras alterações que possam determinar variação do valor patrimonial tributário do prédio; f) Ter-se conhecimento da não inscrição de um prédio na matriz; h) Ser ordenada uma atualização geral das matrizes. **4** - As inscrições ou atualizações matriciais devem referir o ano em que tenham sido efetuadas, bem como os elementos que as justifiquem.

Às câmaras municipais compete (Artigo 128º) colaborar com a administração fiscal no cumprimento do disposto no presente Código, devendo, nomeadamente: a) Remeter ao serviço de finanças competente, até final ao mês seguinte ao da sua aprovação, os alvarás de loteamento, licenças de construção, plantas de arquitetura das construções correspondentes às telas finais, licenças de demolição e de obras, pedidos de vistorias, datas de conclusão de edifícios e seus melhoramentos ou da sua ocupação, bem como todos os elementos necessários à avaliação dos prédios.

Anexo 4 – Relatórios elaborados pelo GC 2021, disponíveis para consulta

- *Estudo de viabilidade para a mudança do modelo censitário em Portugal – relatório de progresso e próximos desenvolvimentos, GC 2021, agosto 2014*



Estudo viabilidade
mudança modelo_Ago

- *Análise dos ficheiros administrativos – Emprego e Mercado de Trabalho, Relatório de progresso, GC 2021, julho 2014*



Análise_EMT_julho20
14.pdf

- *Registo da população e ligação de ficheiros – Relatório de progresso, GC 2021, julho 2014*



Registo pop e ligação
ficheiros_julho2014.pc

- *Análise de ficheiros administrativos – IMI, GC 2021, junho 2014*



AnáliseIMI_junho2014
.pdf

- *Avaliação dos modelos censitários utilizados noutros países e a sua adequabilidade a Portugal, GC 2021, junho 2014*



Avaliação modelos
países_junho2014.pdf

- *Censos 2021 – estudos de viabilidade – cenários para o novo modelo, GC 2021, abril 2014*



Cenário novo
modelo_abril2014.pdf

- *Análise de ficheiros administrativos – Ligações numéricas BDIC/SS/QP/SEF/CGA, GC 2021, dezembro 2013*



Análise_Ligações
numericas_dezembro2

- *Análise dos ficheiros SS e RU/QP para uso nos Censos 2021, GC 2021, dezembro 2013*



AnáliseSS_QP_dezem
bro2013.pdf

- *Análise dos Ficheiros administrativos – BDIC 2010, GC 2021, dezembro 2013*



Análise_BDIC2010.pdf