



REPLAN

Rede de Serviços de Planeamento
e Prospetiva da Administração Pública

O acesso a microdados individuais: situação atual e recomendações

Abril 2026

FICHA TÉCNICA

Título

O acesso a microdados individuais: situação atual e recomendações

Data

Abril 2026

Autoria

Equipa Multissetorial de Monitorização da REPLAN – Rede de Serviços de Planeamento e Prospetiva da Administração Pública

Nota

O presente relatório é da responsabilidade exclusiva da Equipa Multissetorial de Monitorização da REPLAN. O Instituto Nacional de Estatística (INE) participa nesta Equipa da REPLAN com o estatuto de observador e foi um membro ativo nesta atividade pelas apresentações realizadas em outubro de 2025 e que deram início a este trabalho, contudo não é responsável pelos conteúdos formulados, nem pelas recomendações nele constantes.

replan@planapp.gov.pt

Índice

1. Enquadramento.....	4
2. Objetivos	6
3. O acesso a microdados: requerentes e modalidades	7
3.1. Critérios de acreditação dos requerentes.....	8
3.2. Modalidades de acesso	9
4. O processo de acreditação	12
5. Necessidade de tratamento dos dados	15
6. A viabilidade de acesso dos Gabinetes de Estudos	17
7. Conclusões e recomendações.....	19

1. Enquadramento

O Plano de Atividades de 2025 da Equipa Multissetorial de Monitorização (EMM), da REPLAN - Rede de Serviços de Planeamento e Prospetiva da Administração Pública, previa a realização de uma sessão para apresentação das modalidades de acesso a microdados existentes no país, e de análise e debate sobre a viabilidade dessas modalidades serem utilizadas pelos Gabinetes de Estudos, Planeamento, Avaliação e Prospetiva, ou entidades de natureza similar, de forma a colmatar as suas necessidades de informação/evidências para o exercício efetivo das suas funções, no âmbito das respetivas Áreas Governativas.

O acesso a microdados anonimizados em Portugal, nos restantes países da União Europeia e no próprio EUROSTAT, especialmente os dados que resultam de recolhas realizadas no âmbito dos Sistemas Estatísticos Nacionais está regulamentado pelas Leis que regem esses sistemas.

Os Sistemas Estatísticos Nacionais dos países da União Europeia fazem parte do Sistema Estatístico Europeu (SEE), regido por um quadro legislativo (o atual Regulamento (UE) 2024/3018)¹ e um código de boas práticas (Código de Conduta para as Estatísticas Europeias), que garantem a produção de estatísticas fiáveis, harmonizadas e comparáveis, com foco na independência, integridade, qualidade e proteção dos dados. O SEE envolve o Eurostat e os Institutos Nacionais de Estatística (INE) e outras autoridades estatísticas com delegação de competências.

A Lei do Sistema Estatístico Nacional (SEN) em vigor em Portugal é a Lei n.º 22/2008, de 13 de maio, que define a estrutura, os princípios gerais e normas do SEN e da produção de estatísticas oficiais, garantindo independência técnica, a recolha obrigatória de dados para fins estatísticos, a qualidade e a salvaguarda da confidencialidade dos dados.

Em termos de quadro legislativo, a proteção dos dados pessoais foi reforçada pela Lei n.º 58/2019, de 8 de agosto, que assegura na ordem jurídica nacional, o estipulado pelo Regulamento (UE) 2016/679 do Parlamento e do Conselho, de 27 de abril de 2016 (o Regulamento Geral de Proteção de Dados - RGPD), relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados, estabelecendo regras sobre o acesso e a utilização de informações sensíveis.

No que se refere à divulgação e cedência de dados, o Artigo 6.º da Lei do SEN, relativo ao Segredo Estatístico, prevê o seguinte:

- Todos os dados estatísticos individuais recolhidos pelas autoridades estatísticas são de natureza confidencial, pelo que não podem ser divulgados de modo que permitam a identificação direta ou indireta das pessoas singulares e coletivas a que respeitam.
- Os dados estatísticos individuais respeitantes a pessoas singulares não podem ser cedidos, salvo se o seu titular tiver dado o seu consentimento expresso ou mediante autorização do CSE,

¹ Regulamento (UE) 2024/3018 do Parlamento Europeu e do Conselho, de 27 de novembro de 2024, que veio alterar o Regulamento (CE) n.º 223/2009 do Parlamento Europeu e do Conselho, de 11 de março de 2009 e o Regulamento (UE) 2015/759 do Parlamento Europeu e do Conselho, de 29 de abril de 2015.

que delibera, caso a caso, sobre pedidos devidamente fundamentados, quando estejam em causa ponderosas razões de saúde pública, desde que anonimizados e utilizados exclusivamente para fins estatísticos, sob compromisso expreso de absoluto sigilo em relação aos dados fornecidos.

- Os dados estatísticos individuais respeitantes a pessoas coletivas não podem ser cedidos, salvo se os respetivos representantes tiverem dado o seu consentimento expreso ou mediante autorização do CSE, que delibera, caso a caso, sobre pedidos devidamente fundamentados, quando estejam em causa ponderosas razões de saúde pública, planeamento e coordenação económica, relações económicas externas ou proteção do ambiente e desde que sejam utilizados exclusivamente para fins estatísticos, sob compromisso expreso de absoluto sigilo em relação aos dados fornecidos.
- Além dos casos referidos anteriormente, os dados estatísticos individuais sobre pessoas singulares e coletivas só podem ser cedidos para fins científicos, sob forma anonimizada, mediante o estabelecimento de acordo entre a autoridade estatística cedente e a entidade solicitante, no qual são definidas as medidas técnicas e organizativas necessárias para assegurar a proteção dos dados confidenciais e evitar qualquer risco de divulgação ilícita ou de utilização para outros fins aquando da divulgação dos resultados.
- São considerados como visando fins científicos, os pedidos de cedência de dados efetuados no âmbito de um projeto científico concreto, por investigadores de universidades ou de outras instituições de ensino superior legalmente reconhecidas e organizações, instituições ou departamentos de investigação científica reconhecidos pelos serviços competentes.

O acesso e disponibilização de microdados individuais, atualmente em vigor no nosso país, é efetuado tendo por referência este enquadramento jurídico.

2. Objetivos

A sessão “Acesso a microdados - Modalidades e boas práticas” ocorreu no dia 15 de outubro de 2025, pelas 14:30, e teve lugar na Biblioteca do Ministério do Trabalho, Solidariedade e Segurança Social, em Lisboa. Estiveram presentes 40 representantes da quase totalidade dos organismos da EMM.

O primeiro objetivo foi apresentar aos participantes as modalidades de acesso a microdados atualmente disponíveis no país, nomeadamente, de dados alojados na Infraestrutura Nacional de Dados (IND), da responsabilidade do INE e, por sua vez, apresentar um exemplo do necessário tratamento da informação de algumas bases de dados, nomeadamente, os dados da AT sobre o IRS, antes da sua integração nesta infraestrutura, de forma as mesmas poderem ser usadas pelos utilizadores (incluindo os utilizadores internos ao INE e os utilizadores externos). As apresentações estiveram a cargo das duas representantes do INE na EMM².

O segundo objetivo foi promover o debate e reflexão sobre as necessidades específicas de acesso/obtenção de informação estatística por parte dos Gabinetes de Planeamento e entidades similares, para o efetivo exercício das suas funções, e a discussão e análise sobre a viabilidade da utilização das modalidades de acesso existentes por parte daqueles organismos para colmatar as suas necessidades.

Finalmente, o terceiro objetivo foi produzir este relatório com recomendações para a operacionalização de um acesso a microdados individuais anonimizados, que melhor responda às necessidades específicas do exercício de funções destes organismos da Administração Pública, considerando os microdados alojados na IND do INE, recomendações que serão aplicadas também a outras entidades que possam vir a conceder acesso a microdados individuais anonimizados, para os mesmos fins.

² Dra. Ana Margarida Rosa (DI – Serviço de Difusão) – “O Valor Público das Estatísticas Oficiais - Microdados para fins de investigação científica” e Dra. Sofia Rodrigues (Dados Administrativos e Empresariais – Departamento de Recolha e Gestão de Dados) – “Acesso a microdados – O exemplo do IRS”.

3. O acesso a microdados: requerentes e modalidades

Antes de mais importa referir que as modalidades de acesso a microdados que foram objeto da sessão de trabalho se reportavam exclusivamente a dados individuais alojados na IND, tratando-se de dados produzidos pelas Autoridades Estatísticas (o próprio INE, as respetivas delegações do INE das duas regiões autónomas, o Banco de Portugal e as outras entidades com delegação de competências) e de dados administrativos de outros organismos que não são Autoridades Estatísticas, que são transmitidos ao INE através de protocolos de cedência de informação para integração na IND.

A apresentação, intitulada “O Valor Público das Estatísticas Oficiais - Microdados para fins de investigação científica”, teve como foco principal:

- Informar sobre o papel do INE enquanto Autoridade Estatística Nacional com a missão de garantir a independência, segurança e a precisão dos dados estatísticos; de assegurar que os dados são precisos, atualizados e relevantes para o público; de apoiar o desenvolvimento e implementação de políticas nacionais eficazes de gestão de dados; e de disponibilizar informação estatística robusta para suporte à tomada de decisão política.
- Referir a IND enquanto modelo de governança da informação que permite potenciar o acesso a dados para a produção de estatísticas oficiais e para fins de investigação; que aposta na integração dos dados e metadados através de ferramentas e plataformas unificadoras; no garante da segurança e qualidade dos dados, o que significa proteger contra acessos não autorizados e manter altos padrões de integridade e confiabilidade; na aplicação de métodos inovadores de análise avançada, como a Inteligência Artificial, *Machine Learning*, Processamento de Linguagem Natural e *Web Scrapping*, para melhorar o tratamento e análise dos dados.
- Apresentar as modalidades de acesso e disponibilização de microdados aos utilizadores.

O acesso a microdados atualmente em vigor em Portugal é garantido pelo INE, e efetuado de acordo com a Lei do SEN, que define que os dados estatísticos individuais de pessoas singulares e coletivas, salvo raras exceções, apenas podem ser cedidos de forma anonimizada e para fins científicos.

Neste âmbito, foi assinado um protocolo de cooperação³ entre o INE, a Direção-Geral de Estatísticas da Educação e Ciência (DGEEC) e a Fundação para a Ciência e Tecnologia (FCT), com o objetivo de estabelecer as formas de cooperação entre as três entidades no que respeita ao estabelecimento das condições de acesso, para fins de investigação, a dados estatísticos individuais anonimizados constantes de bases residentes no INE, produzidas pelo INE, pelas entidades com delegação de competências e outras fontes, integradas na IND⁴, tendo em vista implementar o disposto nos n.º 7 e

³ Este Protocolo existe desde o início dos anos 2000, com versões que foram atualizadas em 2008, 2014 e 2022 (esta última a que se encontra atualmente em vigor).

⁴ Salvaguardando que os protocolos do INE com as entidades detentoras dos dados permitem a cedência dos mesmos a investigadores, para fins científicos.

n.º 8 do artigo 6.º da Lei do SEN. Neste Protocolo estão definidos quais os critérios de acreditação dos requerentes dos microdados e em que modalidades o acesso pode ser efetivado, cujas especificações serão abordadas nos capítulos seguintes.

3.1. Critérios de acreditação dos requerentes

O acesso aos microdados pode ser concedido a requerentes cujos critérios de elegibilidade são os seguintes:

- Ser investigador em atividade nas entidades, estruturas e redes dedicadas à produção, difusão e transmissão do conhecimento no sistema nacional de ciência e tecnologia (instituições de I&D, laboratórios colaborativos, centros de interface tecnológicos, infraestruturas de ciência e tecnologia, redes e consórcios de ciência e tecnologia – conforme o Decreto-Lei n.º 63/2019 de 16 de maio, artigo 14º e artigo 15º), detentores de identificação e que constem nas bases de dados individuais dos respondentes ao Inquérito ao Potencial Científico e Tecnológico Nacional (IPCTN) ou do Observatório do Emprego Científico e Docente⁵, ambos da responsabilidade da DGEEC.
- Ser estudante com teses de doutoramento em curso e que constem no registo nacional oficial de teses de doutoramento (estabelecido no RENATES, ou outro que o venha a substituir).
- Ser estudante de mestrado, desde que o pedido seja assinado por orientador da tese em curso, docente em instituição de ensino superior nacional legalmente reconhecida.
- Indivíduos que executem atividades de I&D em organismos internacionais de reconhecida reputação e experiência em investigação científica, nomeadamente, agências/estruturas especializadas, do âmbito da Comissão Europeia; agências especializadas das Nações Unidas (Organização Internacional do Trabalho (OIT), Organização para a Alimentação e Agricultura (FAO), Organização das Nações Unidas para a Educação, Ciência e Cultura (UNESCO), Organização Mundial da Saúde (OMS), Grupo Banco Mundial (*World Bank Group*); Fundo Monetário Internacional (IMF); e Organização para a Cooperação e Desenvolvimento Económico (OCDE).
- Ser investigador de outros organismos de investigação que constem na lista de entidades reconhecidas pelo Eurostat na data do acesso aos dados, nos termos do Artigo n.º 4 do Regulamento n.º 557/2013, que aplica o Regulamento n.º 223/2009 do Parlamento Europeu e do Conselho, relativo às estatísticas europeias, no que diz respeito ao acesso a dados confidenciais para fins científicos.

⁵ O Observatório do Emprego Científico e Docente (OECD) é um registo público nominativo que lista todos os contratos de titulares de doutoramento envolvidos em atividades de investigação (I&D) ou de gestão e comunicação de ciência e tecnologia, e contratos de outros trabalhadores integrados em qualquer categoria da carreira de investigação científica, celebrados com a instituição à qual estiveram vinculados. No caso das instituições de Ensino Superior, abrange ainda contratos celebrados com docentes e com pessoal em cargos de gestão (Decreto-Lei n.º 156/2019, de 22 de outubro).

3.2. Modalidades de acesso

Os microdados individuais só podem ser acedidos sob a forma anonimizada, nos termos previstos da alínea d) do artigo 2.º da lei do SEN, e de acordo com as recomendações e melhores práticas de instituições estatísticas internacionais. A **anonimização** implica que os dados são previamente sujeitos a tratamento de confidencialidade antes da sua disponibilização, para garantir que a utilização dos mesmos por parte dos investigadores não permite identificar, direta ou indiretamente, as unidades estatísticas (indivíduos ou empresas). A principal regra de anonimização dos dados é a não disponibilização da identificação direta dos indivíduos ou empresas. Associada a esta existem outros procedimentos, dos quais se salientam: a não disponibilização de variáveis na sua máxima desagregação (por exemplo, a disponibilização apenas ao nível da NUTS II e ao nível da classificação da atividade económica a 2 dígitos); a não disponibilização de valores reais mas de valores médios para indivíduos ou empresas com características idênticas (por exemplo, a média de volume de negócios); ou ainda a transformação de variáveis ordinais em variáveis agregadas em escalões (por exemplo, por escalão etário ou escalão de volume de negócios).

As **modalidades de acesso** aos microdados estatísticos anonimizados são as que se apresentam de seguida:

- A **cedência numa plataforma de cloud dedicada**. Trata-se de uma *cloud* interna do INE onde se disponibilizam os microdados solicitados para projetos acreditados. Este acesso é baseado num sistema de dupla autenticação (*Two-Factor Authentication*), feito de modo a garantir o acesso apenas ao investigador principal.
- Apuramentos realizados pelo INE, mediante pedido específico do investigador, que são disponibilizados sob a forma de **quadros estatísticos de dados**. Esta é uma forma atualmente menos utilizada, mas ainda existem alguns inquéritos mais antigos do INE que não estão disponíveis em bases de microdados.
- Quando a natureza dos dados ou a investigação justificadamente o exija, o acesso pode ser efetuado por apuramento realizado pelo investigador diretamente nas bases de dados individuais, sem identificação direta das unidades estatísticas, ficando o investigador sujeito ao segredo profissional, nos termos da alínea d) do n.º 2 do artigo 6.º da Lei do SEN. Este acesso é realizado em **ambiente de sala segura (safe centre)**, nas instalações do INE, em Lisboa, Coimbra ou Porto e também, mais recentemente, na Direção Regional de Estatísticas da Madeira (DREM). O acesso é efetuado de acordo com as políticas e procedimentos de segurança da informação em vigor no INE, o que implica o estrito controlo pelo INE da informação acedida e uma verificação prévia dos apuramentos efetuados pelo investigador (*output check*), de modo a garantir que estes não contêm dados confidenciais e, desta forma, não permitem a identificação direta e indireta.

Em ambiente de **safe centre** os computadores estão fechados ao exterior, e não é permitida a entrada e/ou uso de qualquer dispositivo eletrónico de dados e/ou imagem, nomeadamente computadores portáteis, tablets, telemóveis, máquinas de filmar e/ou fotografia.

Os apuramentos são realizados através dos softwares que o INE disponibiliza como, por exemplo, o R, STATA ou SPSS, sendo eventualmente, possível disponibilizar outros, a pedido.

Às bases de microdados disponibilizadas em *safe centre* não é feito outro tratamento de confidencialidade além da eliminação do identificador direto, pelo que as mesmas não podem ser disponibilizadas via *cloud*. Trata-se de bases que contém dados muito sensíveis e que são por norma volumosas, pelo que o tratamento de confidencialidade, para a disponibilização em Cloud a que estariam sujeitas iria retirar o interesse da informação e desvirtuar os dados e, conseqüentemente, os resultados das análises sobre os mesmos. São exemplo de bases de dados de acesso exclusivo em *safe centre* os dados administrativos da AT, IMT, IRS, E-fatura, a série de dados do inquérito às rendas de habitação e os dados relativos a óbitos por causa de morte.

Em domínios com risco elevado de reidentificação por combinação de variáveis e cruzamento de fontes (por exemplo, justiça), a proteção não se esgota na anonimização. Deve ser assegurada a avaliação documentada do risco de divulgação e a aplicação de medidas de controlo de divulgação estatística (*statistical disclosure control*), incluindo regras para outputs (por exemplo, limiares mínimos, agregação e supressão) e validação prévia dos resultados (*output checking*), bem como limitação e controlo de exportações.

Todas as bases que são disponibilizadas via *cloud* poderão ser preparadas para acesso em *safe centre*, no caso da granularidade dos dados de origem o permitirem (por exemplo, os dados foram recolhidos/obtidos a nível de freguesia e o tratamento de confidencialidade implica disponibilizá-los só ao nível de NUTS II, ou, os dados existem por CAE a 5 dígitos, mas o tratamento de confidencialidade implica disponibilizá-los só a 2 dígitos da CAE). Se o projeto de investigação justificar o acesso a dados mais desagregados, o investigador pode pedir o acesso às mesmas em ambiente *safe centre*, sujeitando-se às regras inerentes a este tipo de acesso. Na lista das bases de dados abrangidas pelo Protocolo de acesso a microdados individuais estão identificadas aquelas que são de uso exclusivo em *safe centre*. As outras, caso o investigador necessite, terá de contactar o INE para averiguar sobre a possibilidade de consulta a partir de *safe centre*.

- Além das formas de acesso referidas anteriormente e que implicam a acreditação do investigador, o INE prepara ficheiros com informação ao nível da unidade de observação, os designados **Ficheiros de Uso Público** (FUPs), disponíveis no seu Portal, com o objetivo de responder às necessidades dos utilizadores em aceder a informação mais detalhada.

Estes ficheiros (dados e metainformação) contém registos anonimizados, tratados e preparados de forma que a unidade de observação não possa ser identificada, direta ou indiretamente, exceto quando se trate de dados estatísticos individuais sobre a Administração Pública. Estão conforme o princípio do segredo estatístico e de proteção de dados pessoais e são de acesso gratuito a todos os utilizadores que, ainda assim, para acederem aos mesmos têm de aceitar previamente as condições de utilização. Após isso, os ficheiros podem ser descarregados e utilizados nos seus computadores pessoais. Existem atualmente 30 FUPs disponibilizados no site do INE, associados às seguintes temáticas:

Censos; Recenseamento agrícola; Museus públicos; Hospitais públicos e Centros de saúde, alguns dos quais para vários anos.⁶

⁶ Para mais informação sobre a especificidade dos dados disponibilizados nos ficheiros consultar: https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_pufs&menuBOUI=13707095&contexto=up&sefTab=tab5&xlang=pt

4. O processo de acreditação

O processo de acreditação dos investigadores é da responsabilidade da DGEEC. Desde setembro de 2024, o mesmo é efetuado a partir de uma plataforma online⁷, que substituiu o processo até então feito manualmente em versão papel ou em ficheiros eletrónicos em formato PDF.

Antes de proceder ao pedido de acreditação para acesso às bases de microdados, os requerentes têm de criar um perfil de utilizador, no qual são obrigados a colocar a sua identificação e a criar uma senha de acesso para entrada na plataforma de acreditação. A partir do seu perfil, o requerente também pode aceder a um *backoffice* que lhe permite visualizar todos os pedidos que já efetuou, por datas, e o estado em que os mesmos se encontram (em validação, validados, não validados).

A plataforma de acreditação disponibiliza um *frontoffice* que, na sua página inicial, dispõe de informação auxiliar sobre o processo de acreditação, nomeadamente, o Protocolo estabelecido entre o INE, a DGEEC e a FCT(ficheiro para consulta); as etapas, desde a entrada do pedido, até à cedência dos dados; bem como a lista das bases de microdados acessíveis (atualizada regularmente, quando novas bases, ou novos anos, passam a estar disponíveis), organizada por temas e com a respetiva metainformação (a periodicidade, o tipo de dados, os anos disponíveis, o dicionário das variáveis - descrição e descodificação) ou mesmo os próprios instrumentos de notação⁸.

Para formalizar o pedido de acreditação e de cedência de microdados, após o início de sessão, o requerente tem de preencher um formulário. No formulário consta o “Código de conduta”, cujas disposições contemplam, na generalidade, as medidas técnicas e organizativas necessárias para assegurar a proteção dos dados confidenciais e evitar qualquer risco de divulgação ilícita ou de utilização para outros fins aquando da divulgação dos resultados. A subscrição deste Código é parte integrante e obrigatória do pedido de acreditação e de cedência de dados, e terá de ser feita pelo requerente e pelo dirigente do órgão competente da unidade de investigação em cujas atividades se enquadre o projeto de investigação que justifica a cedência dos dados estatísticos individuais. Este Código constitui igualmente parte integrante do acordo entre a autoridade estatística cedente e o requerente e o incumprimento, por ação ou omissão, das normas nele estabelecidas constitui quebra do dever de sigilo (segredo estatístico), podendo incorrer nas sanções previstas nos artigos 25.º a 33.º da Lei SEN. Adicionalmente, também constam uma Declaração de compromisso” e um “Aviso de privacidade e consentimento”, que obrigam ambos assinalar a tomada de conhecimento.

A informação requerida no formulário é, na sua quase totalidade, de preenchimento obrigatório, salvo nas situações em que não se aplique, conforme discriminada abaixo:

- Dados do **requerente** - Nome completo, Número de identificação, Correio eletrónico, Telefone e CencialD (nos casos em que se aplique).

⁷ <https://acreditacao.dgeec.medu.pt/>

⁸ https://www.ine.pt/xportal/xmain?xpid=INE&xpqid=ine_bd_inv_fct&ine_smenu.boui=161302718&ine_smenu.selected=161302852

- Identificação de **outros investigadores** que terão também acesso aos dados - Nome completo, Número de identificação, Correio eletrónico e CienicalD (nos casos em que se aplique). Esta informação apenas deve ser preenchida nos casos em que outros investigadores estão integrados no projeto e vão aceder aos dados. Até à data de fim do projeto, podem ser adicionados (a pedido do requerente) outros investigadores não contemplados inicialmente, aplicando-se os mesmos procedimentos necessários.
- Identificação das **bases de microdados** – Nome da base e indicação do período de referência dos dados. A escolha é baseada na lista disponibilizada pelo INE, e os campos são preenchidos a partir da mesma, sem possibilidade de serem editados.
- Identificação do **projeto de investigação** – Título, Data de início e Data de fim (esta última pode ser prolongada, a pedido do requerente), Objetivos, Plano de trabalhos, Integração em programas de formação (caso se aplique), Nome da instituição onde se insere o projeto, Nome e Cargo do responsável pela instituição e áreas de investigação do projeto.

Após o preenchimento de toda a informação do formulário, é necessário gerar um ficheiro eletrónico em formato PDF, que deve ser assinado (digital ou manualmente) pelo requerente dos dados e pelo responsável da instituição onde se insere o projeto. Para além do requerente, cada um dos outros investigadores que irão ter acesso aos dados, e que foram incluídos no pedido, devem obrigatoriamente assinar uma declaração de compromisso e um aviso de privacidade e consentimento. Para tal, após o preenchimento da informação individual por investigador, o requerente tem de gerar ficheiro(s) eletrónico(s), também em formato PDF, que devem ser descarregados para serem assinados pelos próprios antes de voltarem a ser carregados e anexados ao pedido geral de acreditação.

Só depois de todos os documentos preenchidos, assinados e carregados na plataforma, o requerente pode submeter o pedido de acreditação que, seguidamente, será avaliado pela DGEEC e, cumprindo todos os requisitos necessários, será validado ou, havendo alguma inconsistência na informação, não será validado. Na situação de pedido validado, o requerente e o INE recebem um aviso automático e a partir desse momento o INE dará continuidade ao processo procedendo à disponibilização dos microdados de acordo com a modalidade adequada. Na situação de pedido não validado, o requerente também recebe um aviso automático, com a justificação da não validação, sendo que, em grande parte das situações, é possível proceder à reabertura do formulário para correção de informação e nova submissão do pedido.

A aprovação de um pedido de acreditação, depois de submetido com toda a informação corretamente preenchida e sem falta de documentos, é bastante rápida, não demorando mais do que um dia, na grande maioria dos casos.

Outra informação relevante que importa referir é que a acreditação é válida pelo período de duração do projeto de investigação, podendo haver extensão do prazo, a pedido do requerente, como já referido anteriormente.

Para cada projeto é necessário proceder a um pedido de acreditação, o que significa que as bases de microdados listadas em determinado pedido só podem ser usadas no âmbito do projeto a que se reporta o pedido.

Através deste processo não se acreditam instituições, mas indivíduos (investigadores), embora estes tenham de estar integrados em instituições que desenvolvam projetos de investigação e seja necessária a assinatura do dirigente da instituição, o que serve para oficializar a institucionalização do projeto. Este processo de acreditação de indivíduos é diferente do processo do EUROSTAT que, em vez de indivíduos, acredita entidades, ou seja, reconhece instituições como entidades de investigação, após um processo de candidatura e de avaliação das mesmas. Os investigadores dessas entidades podem depois submeter os seus pedidos para acesso a microdados, que são avaliados pelo EUROSTAT e objeto de parecer por parte das autoridades estatísticas nacionais que transmitiram as bases de microdados solicitadas no pedido.

5. Necessidade de tratamento dos dados

Antes de passar à análise da viabilidade da utilização destas vias de acesso por parte dos Gabinetes de Estudos e entidades com funções similares, na apresentação intitulada “O acesso a microdados – o caso do IRS”, o INE procurou mostrar a necessidade de um trabalho estruturado e consistente sobre algumas bases de dados, que resultam de fontes administrativas. Esse trabalho consiste num conjunto de operações técnicas e metodológicas destinadas a harmonizar conceitos e classificações, integrar informação de diferentes fontes (internas e externas), e melhorar a qualidade, consistência e compreensão dos dados, de modo a proceder à sua integração na Infraestrutura Nacional de Dados (IND), seja para fins estatísticos (como seja, o próprio INE), seja para fins de investigação (INE e utilizadores externos).

O exemplo apresentado foi o trabalho efetuado aos fluxos de dados relativos ao Modelo 3 do Imposto sobre o Rendimento das Pessoas Singulares (IRS), da Autoridade Tributária (AT), que exigiu um processo de tratamento e estruturação da informação bastante complexo e demorado, não apenas pelos desafios que colocou na integração dos dados, como também nos tratamentos aplicados e descodificação das variáveis recebidas e criação de novas.

No contexto do desenvolvimento da IND, o INE implementou alterações na sua organização de modo a fortalecer a capacidade de gestão e de análise de dados. Esta reestruturação tem vindo a possibilitar a transição de um modelo de tratamento de microdados por projeto (ou operação estatística) para um tratamento centralizado e integrado das fontes de dados. O objetivo passa por criar uma base de dados única, por cada fonte de dados, passível de satisfazer as necessidades dos vários utilizadores, tanto internos (por exemplo, estatísticas dos rendimentos e contas nacionais) como externos (por exemplo, para efeitos de investigação).

Este modelo de processamento de dados foi aplicado ao IRS. Neste caso, no contexto de um protocolo estabelecido entre as duas entidades, a AT transmite ao INE as tabelas dos vários anexos do IRS (num total de 23), em duas versões por ano, a primeira em setembro, para que os dados da AT ainda possam beneficiar as estatísticas que são produzidas até dezembro do ano, e depois em março quando a AT publica o seu dossier estatístico, sendo esta considerada a base de dados final.

As tabelas disponibilizadas pela AT não incluem o ID da declaração, tornando o processo de definição dos agregados fiscais muito desafiante. De modo a poder tirar o melhor partido da informação recebida, nomeadamente para benefício das estatísticas de rendimentos das famílias, o INE definiu um conjunto de procedimentos, tendo em conta não somente os dados recebidos, mas também informação proveniente de outras fontes existentes no INE. De uma forma geral, o processo de construção dos agregados fiscais passou pelo agrupamento dos vários sujeitos passivos e dependentes conforme as posições e descrições que constam nos modelos do IRS (ex.: sujeito passivo A/B, dependente, dependente deficiente, guarda conjunta). Definiu-se que o NIF classificado como sujeito passivo A era o cabeça do agregado fiscal e os outros NIF classificados como B, D, DD ou DG eram, respetivamente,

o cônjuge ou unido de facto com quem entrega a declaração, os dependentes, os dependentes deficientes e os dependentes em guarda conjunta.

Este processo de tratamento da informação implicou a comparação e validação dos dados com outras fontes, de forma a afinar os pressupostos e a metodologia usada. Neste caso em concreto, foram feitas validações entre os resultados dos dados tratados e os resultados do dossier fiscal da AT, assim como, por exemplo, comparações com os valores da base da população residente, para verificar se havia coerência entre os dados, ou necessidade de fazer alguns ajustes nos pressupostos.

A criação do agregado fiscal permitiu ao INE integrar a informação do IRS nas suas operações estatísticas. A integração com os dados da base da população residente, permitiu enriquecer os dados do IRS com as variáveis demográficas com elevado grau de fiabilidade, como o sexo e a data de nascimento, o que, na definição do agregado fiscal, ajudou a tomar decisões relativamente aos dependentes de guardas conjuntas.

A ligação ao ICOR, o Inquérito às condições de vida e rendimento, permitiu que os dados do IRS passassem a ser integrados nos apuramentos sobre a distribuição de rendimento proveniente de trabalho dependente, independente e pensões. Foi igualmente possível produzir outros estudos, nomeadamente o impacto do nascimento do primeiro filho no orçamento familiar, publicado na área StatsLab, disponível no Portal do INE.

No contexto da passagem para um processo de tratamento centralizado das fontes externas, o INE passou a criar fichas de caracterização mais detalhadas destas bases de microdados, incluindo as tabelas de descodificação e todas as variáveis disponibilizadas, descrevendo os principais pressupostos usados para determinar os agregados fiscais. À data, foi possível criar estas fichas detalhadas para os dados do IRS e do e-Fatura e são disponibilizadas para consulta dos utilizadores aquando do pedido de acreditação.

A apresentação do exemplo do IRS ilustra a complexidade do tratamento dos microdados, a necessidade de correção de duplicações e transições de posição e a relevância da integração dos dados e centralização dos tratamentos efetuados para a produção de estatísticas públicas fiáveis.

6. A viabilidade de acesso dos Gabinetes de Estudos

O enquadramento, pressupostos e procedimentos relativos ao acesso a bases de microdados individuais anonimizados que foram apresentados na sessão de trabalho e descritos nos capítulos anteriores, levaram a debater sobre a sua viabilidade e adequabilidade para responder às necessidades de informação/evidências dos Gabinetes de Estudo, Avaliação, Prospetiva e Planeamento e de outras entidades com funções similares. Na sessão de debate várias preocupações foram manifestadas pelos participantes dos organismos com representação na EMM, nomeadamente, a de elucidar como é que estas entidades da Administração Pública, podem aceder aos dados, uma vez que não são entidades de investigação e podem não ter indivíduos que se encontrem a realizar doutoramentos ou mestrados.

Embora algumas entidades em causa façam parte do diretório de inquirição do IPCTN (ver especificidades do inquérito na caixa de informação da página seguinte), que é uma das principais fontes de acreditação dos investigadores, conjuntamente com o registo público nominativo do Observatório do Emprego Científico e Docente, outras têm dificuldade em reconhecer que as suas funções e atividades de monitorização, avaliação e prospetiva, de suporte à tomada de decisão e definição de políticas públicas, se enquadram nos conceitos de Investigação e Desenvolvimento (I&D). Neste sentido, será necessário equacionar outro tipo de enquadramento e, se necessário, **um enquadramento legal específico** que, contemplando a natureza das suas funções, possibilite o acesso aos microdados individuais a estes organismos da Administração Pública, em particular.

Outras preocupações manifestadas prendem-se com as formas e as modalidades de acesso aos dados, sobretudo às bases que apenas podem ser consultadas em *safe centre*. As regras de funcionamento do *safe centre* acabam por se repercutir em grandes constrangimentos para estes organismos, que têm que se deslocar às instituições onde os mesmos estão instalados, não podem usar os equipamentos de comunicação, o que impede contactos entre as equipas de trabalho, não podem utilizar os computadores pessoais para efetuar os apuramentos, os *outputs* dos resultados têm de ser previamente submetidos a processos de *output checking* antes de ser entregues, o que significa um tempo de espera para os receberem. Adicionalmente, o facto do acesso ao *safe centre* ser condicionado por processos de agendamento prévios, tudo isto são constrangimentos que dificultam uma resposta rápida às necessidades de informação, imediatas, urgentes e recorrentes, destes organismos no âmbito do exercício das suas funções.

As estatísticas oficiais sobre I&D em Portugal são produzidas a partir do **Inquérito ao Potencial Científico e Tecnológico Nacional (IPCTN)**, que integra o Sistema Estatístico Nacional (SEN) e recolhe informação sobre recursos humanos e despesa em I&D, de entidades enquadradas em quatro setores de execução: Empresas, Estado, Ensino Superior e Instituições Privadas sem Fins Lucrativos (IPSFL).

Os resultados desta operação estatística possibilitam a construção dos indicadores de caracterização e evolução do sistema científico e tecnológico nacional e integram as séries estatísticas oficiais nacionais (INE) e internacionais (OCDE e EUROSTAT, entre outros). Este inquérito é realizado em conformidade com critérios definidos a nível internacional pelo EUROSTAT, em articulação com a OCDE, e tem como referência o Manual de Frascati (2015).

Os conceitos, classificações e metodologias de inquirição são discutidas a nível internacional em grupos de trabalho de peritos no âmbito daquelas organizações, nos quais Portugal faz parte, resultando nas recomendações inscritas no Manual de Frascati e em outros documentos auxiliares, que permitem garantir uma maior comparabilidade internacional destas estatísticas.

O conceito de I&D é abrangente e inclui a “Investigação Fundamental (IF)” e “Investigação Aplicada (IA)”, mais relacionadas com a investigação realizada em meios académicos e em outras instituições científicas, mas também considera o “Desenvolvimento experimental (DE)”, cujas atividades estarão mais próximas da investigação realizada no setor empresarial e em alguns organismos do Estado.

Para elucidar os respondentes sobre se as atividades/projetos que realizam, no seu dia a dia, podem ser consideradas atividades de I&D, são apresentados no inquérito alguns exemplos, organizados por áreas, dos quais exemplificamos os abaixo, por parecerem mais relevantes no âmbito da temática deste relatório.

Na área da **Economia e Gestão**: trabalhos sobre a revisão das teorias dos fatores que determinam as desigualdades regionais no crescimento económico; estudos abstratos sobre teoria económica, que se concentra em descobrir se há equilíbrio natural numa economia de mercado; e o desenvolvimento de novas teorias do risco (são exemplos de IF). A análise de um caso regional específico com o objetivo de desenvolver políticas governamentais; a investigação sobre as propriedades de um mecanismo de leilão que pode ser relevante para o espectro das telecomunicações; a investigação de novos tipos de seguros para cobrir novos riscos de mercado ou novos tipos de instrumentos de poupança (são exemplos de IA). O desenvolvimento de modelos operacionais, baseados em resultados estatísticos, para conceber instrumentos de política económica que permitam a uma região recuperar o atraso em questões de crescimento; e o desenvolvimento de um novo método de gestão dos fundos de investimento (são exemplos de DE, desde que existam provas de inovação).

Na área da **Educação**: a análise de fatores ambientais que influenciam a capacidade de aprendizagem; e o estudo do efeito de diferentes tipos de material pedagógico na forma como os alunos aprendem estratégia matemática, modificando o material pedagógico e medindo o que os alunos aprenderam através de instrumentos padronizados (são exemplos de IF). A avaliação comparativa dos programas nacionais de educação destinados a colmatar o fosso de aprendizagem das comunidades desfavorecidas e o estudo para a implementação de um currículo específico de matemática para determinar o que os professores precisam para implementá-lo com sucesso (são exemplos de IA). O desenvolvimento de testes para selecionar programas educacionais que devem ser usados para crianças com necessidades especiais e o desenvolvimento e verificação (em sala de aula) de software e ferramentas de apoio, com base em trabalho de campo, para melhorar os conhecimentos matemáticos dos alunos de educação especial (são exemplos de DE).

7. Conclusões e recomendações

Os constrangimentos referidos no capítulo anterior resultaram num conjunto de recomendações que a seguir se descrevem:

- Dar a conhecer ao Conselho Superior de Estatística (CSE) as preocupações e dificuldades que estes Organismos da Administração Pública sentem no que respeita ao acesso a dados, dando conta da desadequação do enquadramento e dos procedimentos atualmente em vigor que possibilitam o acesso de microdados apenas para fins de investigação. Para tal foi proposto, no Plano de Atividade de 2026 do CSE, fazer a “Apresentação do Relatório sobre o Acesso a microdados - Equipa Multissetorial de Monitorização”, numa das reuniões da Secção Permanente de Coordenação Estatística.
- Aproveitando a retoma do processo de revisão da Lei do SEN, propor, em sede própria, que no que concerne ao acesso a microdados, além da cedência e consulta para fins de investigação, se equacione a possibilidade de se contemplar também o acesso para fins de avaliação, monitorização, prospetiva e planeamento, isto é, atividades de apoio à tomada de decisão sobre políticas públicas, quando as mesmas forem realizadas por organismos da Administração Pública que tenham essas competências e funções, previstas nas sua leis orgânicas ou equivalentes.
- A operacionalização do acesso para fins de monitorização/avaliação/prospetiva/planeamento de políticas públicas deve assentar em critérios objetivos e verificáveis, designadamente: existência de mandato institucional e/ou plano de avaliação ou instrumento de planeamento aprovado; delimitação do objeto, variáveis estritamente necessárias (*need-to-know*), *outputs* previstos e prazos; e proibição de utilização para finalidades não declaradas ou incompatíveis.
- Para salvaguarda do princípio da confiança no sistema estatístico, deve ser garantido, preferencialmente em previsão legal expressa, que os microdados cedidos para fins estatísticos e/ou de avaliação/monitorização não podem ser utilizados, direta ou indiretamente, para finalidades administrativas, fiscais, inspetivas, de fiscalização ou sancionatórias, nem ser disponibilizados a entidades (ou unidades orgânicas) com essas competências, assegurando critérios de elegibilidade rigorosos e separação funcional e orgânica relativamente a essas funções.
- Que a proposta anterior se circunscreva aos organismos que fazem parte da Comissão Plenária e/ou das Equipas Multissetoriais da REPLAN, rede interministerial onde estão representadas todas as áreas governativas, coordenada pelo Centro de Planeamento e de Avaliação de Políticas Públicas (PLANAPP), assim como às entidades que lhe venham a suceder no decurso de processos de reestruturação em que as mesmas estejam envolvidas.
- Em alternativa à impossibilidade de considerar o proposto nos pontos anteriores referentes à Lei do SEN, equacionar a necessidade de produzir um **enquadramento legal específico** que possibilite o acesso aos microdados individuais, sobre pessoas singulares e coletivas, anonimizados por parte destes organismos da Administração Pública em particular, atendendo à

especificidade da natureza das suas funções e necessidades de informação, nomeadamente referenciando na legislação orgânica destes organismos.

- Considerando que as necessidades de dados destes Gabinetes são recorrentes e urgentes, recomenda-se que, sobretudo no que respeita aos microdados que só podem ser disponibilizados em ambiente seguro, se procure agilizar esta forma de acesso, procurando uma solução mais prática, designadamente, o acesso aos dados através de *remote safe centres*, a partir dos postos de trabalho individuais. O acesso a bases de dados de consulta exclusiva em ambiente seguro, através de soluções de *remote safe centre* (por exemplo, num modelo semelhante ao BPLIM – Laboratório de Investigação em Microdados, do Banco de Portugal), traria múltiplos benefícios do ponto de vista da operacionalidade e da logística, contribuindo igualmente para promover a digitalização e a cultura de avaliação de políticas públicas na administração pública.

No caso de *remote safe centres*, devem ser definidos requisitos mínimos de segurança, controlo e auditoria, incluindo: autenticação forte (MFA); perfis e segregação de acessos; *logging* integral e rastreabilidade; revisão periódica de acessos; controlo de exportação; *output checking* obrigatório; auditoria; e regime de consequências (suspensão/cessação de acesso e reporte de incidentes).

Em particular para microdados sensíveis, devem ser estabelecidas regras de controlo de divulgação (*statistical disclosure control*) e validação prévia dos *outputs*, com medidas proporcionais ao risco de reidentificação, assegurando que a divulgação de resultados não permite identificação direta ou indireta.

- Em alternativa aos *remote safe centres*, equacionar a possibilidade de uma descentralização maior dos pontos de acesso físico, com:
 - a possibilidade de um deles ficar localizado no perímetro da REPLAN para aí poderem ser acedidos pelos Gabinetes e entidades similares que integram esta rede;
 - adicionalmente, a possibilidade de ficarem também alojados em outras entidades com responsabilidades na avaliação e outras funções de apoio à tomada de decisão sobre políticas públicas que, pelo volume e frequência de pedidos de acesso a microdados que dirigem ao INE, poderão justificar a criação de postos de acesso físico dirigidos exclusivamente aos seus próprios colaboradores⁹.
- Existe a necessidade de trabalhar e disponibilizar metainformação associada às diferentes bases de dados, sobretudo as de fontes administrativas, por forma a promover a harmonização de conceitos, o conhecimento comum das fontes dos dados, a descodificação das variáveis e a interpretação uniformizada das variáveis e dos dados. Além disso, existem solicitações de informação que implicam a necessidade de ligação de inúmeras bases de dados, e esse trabalho, complexo e moroso, obriga a um trabalho conjunto não sendo desejável obter resultados que não

⁹ Sujeitos aos requisitos de segurança, controlo, auditoria e autenticação exigidos pelo INE, bem como, às regras de controlo de divulgação estabelecidas para dados desta natureza.

façam sentido pelo facto de se utilizarem conceitos ou interpretações diferentes entre bases de microdados. Para tal, recomenda-se um trabalho conjunto entre o INE e os membros da REPLAN, para chegar a resultados que façam sentido para todos.

- Deve ser clarificado o modelo de governação do processo, identificando responsabilidades (aprovação de projetos e perfis de acesso; validação de outputs; manutenção de dicionários/metadados e regras de qualidade; versionamento e registos de alterações; periodicidade de revisão; e canal de esclarecimento metodológico), em articulação com o INE e a rede REPLAN.
- Recomenda-se uma implementação faseada do modelo, com monitorização e avaliação contínua (incluindo métricas de uso, tempos de resposta, resultados de auditorias, incidentes e reavaliação periódica do risco de reidentificação), ajustando salvaguardas e procedimentos em função da evidência recolhida.
- Quando existam funções mistas na mesma entidade, o acesso deve limitar-se a unidades com separação orgânica e funcional demonstrável ("*firewalls*"), garantindo a impossibilidade de utilização ou partilha interna de microdados para fins estatísticos e de monitorização/avaliação/prospetiva/planeamento de políticas públicas.
- Recomenda-se a criação de um registo central de projetos de acesso a microdados (entidade requerente, finalidade, bases/variáveis, período, equipa autorizada, outputs autorizados e data de expiração), sujeito a revisão periódica.
- Devem ser definidos prazos de expiração do acesso por projeto e regras claras de retenção e destruição de derivados/outputs intermédios, com verificação.
- Finalmente, não obstante as recomendações feitas, no sentido de acomodar novas necessidades, recomenda-se que qualquer enquadramento de alargamento de acesso a dados individuais (ainda que anonimizados) preveja uma proibição legal expressa de disponibilização ou acesso por entidades, organismos ou unidades orgânicas com competências de inspeção, fiscalização ou aplicação de sanções, bem como critérios de elegibilidade e salvaguardas de separação funcional e orgânica relativamente a essas funções. Esta recomendação tem como preocupação continuar a acautelar o princípio da confiança no Sistema Estatístico Nacional, enquanto ativo estratégico que não deve ser comprometido em qualquer modelo de alargamento de acesso a microdados. A preservação desta confiança depende de uma garantia inequívoca de que os dados recolhidos para fins estatísticos e para fins de avaliação/monitorização de Políticas Públicas não serão utilizados para finalidades administrativas, fiscais, inspetivas, de fiscalização ou sancionatórias.



REPLAN

Rede de Serviços de Planeamento
e Prospetiva da Administração Pública

