

# ACESSO A BASES DE MICRODADOS: APLICAÇÕES E IMPACTOS NA PESQUISA EM CIÊNCIAS SOCIAIS

*Susana da Cruz Martins<sup>1</sup>, Rosário Mauritti<sup>2</sup>  
e António Firmino da Costa<sup>3</sup>*

## RESUMO

A utilização de microdados institucionais para fins de pesquisa tem sido crescente, com aplicações cada vez mais diversificadas e complexas. O uso deste tipo de dados tem auxiliado na fundamentação e explicação de alguns dos processos sociais mais relevantes da contemporaneidade. Neste âmbito, e decorridos mais de 10 anos desde que foi celebrado um Protocolo, em Portugal, que facultou essa possibilidade, analisam-se agora as oportunidades analíticas, os procedimentos e os códigos de conduta que enquadram o acesso dos pesquisadores a estes recursos de informação empírica. Estão aqui em causa as possibilidades de aprofundamento e ampliação deste tipo de acordo a outros contextos, bem como o desenvolvimento de dispositivos vantajosos para a comparação internacional.

**Palavras-chave:** Microdados. Metainformação. Anonimização. Harmonização.

---

<sup>1</sup> Instituto Universitário de Lisboa (ISCTE-IUL), Centro de Investigação e Estudos de Sociologia (CIES-IUL), Portugal. Susana.Martins@iscte.pt

<sup>2</sup> Instituto Universitário de Lisboa (ISCTE-IUL), Centro de Investigação e Estudos de Sociologia (CIES-IUL), Portugal. rosario.mauritti@iscte.pt

<sup>3</sup> Instituto Universitário de Lisboa (ISCTE-IUL), Centro de Investigação e Estudos de Sociologia (CIES-IUL), Portugal. antonio.costa@iscte.pt

## **ACCESS TO MICRO DATASE: APPLICATIONS AND IMPACTS ON RESEARCH IN SOCIAL SCIENCES**

### **ABSTRACT**

The use of institutional micro data for research purposes has been growing, with increasingly diverse and complex applications. The use of this type of data has aided the foundation and explanation of some of the most relevant social processes of contemporaneity. In this context, and after more than 10 years since a Protocol was signed in Portugal, which provided this possibility, this paper analyzes the analytical opportunities, procedures and codes of conduct that frame the access of researchers to these resources of empirical information. We are concerned with the possibilities of the deepening and broadening of this kind of agreement to other contexts, as well as the development of useful devices for international comparison.

**Keywords:** Microdata. Metadata. Anonymization. Harmonization.

### **AS CIÊNCIAS SOCIAIS E O USO DE DADOS ESTATÍSTICOS INSTITUCIONAIS**

A produção das ciências sociais, traduzida em indicadores científicos, tem registado um incremento notável e sustentável nos últimos anos. Em Portugal, e apesar desta área representar apenas 5% das publicações científicas (quadro 1), foi uma das que mais cresceu nos últimos 10 anos, verificável no aumento relativo de 257%. Este tipo de indicador dá conta de uma expansão e profissionalização consistentes desta área científica.

Este movimento enquadra-se noutros mais gerais sobre a importância da modernização e profissionalização do campo científico, Portugal é um dos países onde a taxa de publicações mais cresceu entre 2000 e 2010 (159%) no conjunto de países da UE durante este período (GPEAR, 2010).

### Quadro 1 - Produção científica portuguesa: 1990-2010 - Publicações por área científica

Publicações	Total em milhares 2010	Total em milhares 1990-2010	% em 2010	Taxa (%) crescimento 2000-2010
<b>[1] Publicações Classificadas</b>	12424	113452	—	—
Ciências Exatas	3355	35381	26,1	126
Ciências Naturais	2395	22258	18,6	199
Ciências da Engenharia e Tecnologias	1938	21336	15,1	115
Ciências Médicas e da Saúde	3423	24392	26,6	298
Ciências Agrárias	486	4500	3,8	155
Ciências Sociais	657	4428	5,1	257
Humanidades	134	1028	1,0	159
Multidisciplinares	36	129	0,3	—
<b>[2] Publicações não Classificadas</b>	441	995	3,4	—
<b>Total de Publicações [1 + 2]</b>	12865	114447	100,0	—

Fonte: GPEARI (2010).

Todo este processo de transformação das ciências sociais é atravessado por um conjunto de fatores mobilizadores de produção científica. Entre os quais se destaca a utilização de estatísticas oficiais cada vez mais detalhadas e com uma periodicidade também mais estabilizada. É nesta senda que o reforço da regularidade e o alargamento do acesso tem colocado os indicadores estatísticos como um recurso primordial na produção científica dessas áreas. O seu uso constitui uma verdadeira possibilidade, tanto na pesquisa *fundamental*, com um potencial de aprofundamento e inovação do ponto de vista teórico, metodológico e até analítico; como na de caráter mais *aplicado*, orientada para o apoio a processos de planeamento, de decisão e de avaliação de políticas públicas.

Os usos e aplicações da estatística institucional têm hoje uma presença social de largo espectro. Tal é verificável nos sistemas político-administrativo, económico-empresarial, educacional-formativo ou comunicacional-mediático, repercutindo-se ainda ao nível do quotidiano, na ação coletiva e nos direitos, liberdades e garantias (MAURITTI et al., 2002). A solicitação de informação estatística tem, pois, aumentado de forma muito exponencial e para aplicações muito diversas. Isto ao mesmo tempo que aumenta a necessidade de recolocar alguns dos objetivos de pesquisa e procedimentos de análise em patamares com outros níveis de complexidade.

É nesta senda que no ano 1999, em Portugal, é celebrado um primeiro Protocolo entre o Ministério da Ciência e da Tecnologia (MCT) e o Instituto Nacional de Estatística

(INE), por meio do qual passa a ser permitido o acesso regulamentado a microdados anonimizados por pesquisadores creditados, desde que salvaguardados os princípios fundamentais da *confidencialidade* e segredo estatísticos.<sup>4</sup> No novo quadro institucional de cooperação estratégica, entre organismos produtores de informação estatística e pesquisadores, é criada uma nova configuração de possibilidades para ambas as partes. Do lado dos serviços de estatística, o Protocolo materializa o compromisso de parceria e satisfação das necessidades dos utilizadores, reconhecendo a dupla necessidade de disponibilizar essa informação para uso de pesquisa científica e, em simultâneo, regulamentar a sua utilização. Uma segunda entidade, o MCT fica com a incumbência de credenciar as instituições de pesquisa e pesquisadores que solicitam o acesso. Os pesquisadores, por seu lado, ficam obrigados a garantir requisitos de qualidade, confidencialidade, imparcialidade e objetividade do conhecimento. Entre as responsabilidades individualizadas que assumem estão: não procurar identificar indivíduos ou instituições envolvidos nas bases de dados; não comercializar ou ceder dados a terceiros; referenciar o INE em toda a publicação ou comunicação baseada nesses dados e informá-lo de quaisquer problemas de natureza técnica, metodológica, de registo de dados ou de erros detetados nas bases de dados utilizadas; etc.

### **BASES DE MICRODADOS: QUE POSSIBILIDADES?**

Qual o significado exato de *bases de microdados*? O que as distingue de *macrodados* ou *dados agregados*? A resposta é relativamente simples. Os diversos organismos produtores de estatísticas, assim como entidades que coletam dados administrativos, realizam inquéritos, censitários ou amostrais, reunindo informações sobre um conjunto de unidades de observação. Estas podem ser de vários tipos, desde indivíduos, agregados familiares, casas, empresas, escolas, hospitais, etc. As bases de microdados são os arquivos de dados eletrónicos que contêm informações parametrizadas sobre cada unidade de observação (DUPRIEZ; BOYKO, 2010). Em geral, são organizados em bases de dados onde cada *linha* (ou registo) contém informações (organizadas pelas várias variáveis) sobre cada unidade de observação.

A utilização de microdados institucionais por parte de equipas de pesquisa impulsionou o estabelecimento de relações muito vantajosas no levantamento de algumas das questões mais sensíveis nesses processos. Estas puderam contribuir para o desenvolvimento de *procedimentos técnicos* (suportes tecnológicos e processamento informático das bases de dados); *metainformação estatística* (sistematização, atualização e aperfeiçoamento permanentes, possíveis através de uma prática de

---

<sup>4</sup> Este acordo é mais tarde consolidado com a Lei do Sistema Estatístico Nacional (Lei n.º 22/2008, de 13 de Maio) (PORTUGAL, 2008) e a assinatura de novo Protocolo no mesmo ano.

utilização na pesquisa); *processos de anonimização* (soluções rigorosas e expeditas quanto a garantias de anonimato e confidencialidade, mas com claros benefícios no que respeita a aprofundamentos analíticos); *controle dos dados e resultados a divulgar* (permitindo maiores certezas na sua validade, representatividade, fiabilidade e significância); *comparabilidade* (entre indicadores nacionais e internacionais e entre fontes estatísticas); *aperfeiçoamento* das operações estatísticas e aproveitamento analítico das suas potencialidades (testes e sugestões a partir das experiências de investigação); e *processos de trabalho* (entre técnicos do INE e pesquisadores) (MAURITTI et al., 2002).

O Protocolo alargou, em Portugal, o tipo de informação que os pesquisadores tinham ao seu dispor – que nos seus formatos mais convencionais abrangiam as publicações e relatórios disponíveis ao público em geral; a informação divulgada em linha na Internet; o pedido de dados específicos, não publicados, cedidos já num formato de *output* –, sendo as estatísticas divulgadas nestes três modelos facultadas através de tabelas de frequências ou cruzamentos simples de indicadores agregados. Apesar de se tratar de indicadores construídos, a partir de procedimentos standardizados, comuns aos diversos organismos nacionais e internacionais, e portanto, terem a grande vantagem de permitirem a realização de comparações entre vários momentos e entre países, muitas vezes a sua utilização torna-se insuficiente para o desenvolvimento, de ensaios de categorizações e agregações alternativas, de construção e teste de novas variáveis derivadas, e de análises multivariadas de diversos tipos. Tais realizações dependem de orientações teórico-conceituais complexas ou específicas, por um lado, e, por outro, de um diálogo experimental entre sucessivas variantes operatórias e análises dos respetivos resultados.<sup>5</sup> Estes estudos confirmam que, de fato, alguns dos problemas de pesquisa mais interessantes e que mais podem fazer avançar o conhecimento científico sobre as configurações de sociedade e as suas dinâmicas sociais contemporâneas, requerem a utilização de microdados institucionalmente produzidos.

No plano científico, o trabalho de investigação conduzido com o acesso direto aos microdados tem, pois, as seguintes vantagens: realização de *análises de dados mais finas*; a construção de *variáveis compostas de tipos diversificados*, adaptadas a parâmetros teóricos e temas de investigação; desenvolvimento de *análises multivariadas*, estabelecendo relações entre um grande número de indicadores; e a prática de relação experimental com as bases de dados, conduzindo a processos de interação entre parâmetros teóricos, ensaios de operacionalização e tentativas de interpretação de resultados (MAURITTI et al., 2002). Para além destas existem outras de carácter operacional e de racionalização dos processos e custos da pesquisa (DUPRIEZ;

---

<sup>5</sup> Cf. resultados analíticos da pesquisa desenvolvida pelos autores do texto – que aliás representam uma das equipas pioneiras no uso do Protocolo (MAURITTI, 2004; MARTINS; MAURITTI; COSTA, 2007).

BOYKO, 2010) que implicam na *redução da repetição de coletas de dados*, evitando que as equipes repliquem a produção do mesmo tipo de informação, diminuindo alguma da carga de trabalho associada a este tipos de operações; *racionalizando os custos* da pesquisa no que respeita não só à coleta mas também as formas de disseminação.

### **ANONIMIZAÇÃO, VALIDAÇÃO E METAINFORMAÇÃO: DAS PREOCUPAÇÕES ÉTICAS ÀS PREOCUPAÇÕES TÉCNICAS**

Em toda a política de divulgação de dados estatísticos, e reforçadamente para as bases de microdados, existe uma tensão ética clara entre servir os pesquisadores de um recurso, de grande valor para análises complexas e avançadas, e garantir que alguns dos princípios deontológicos mais relevantes por parte de quem produz e divulga informação sejam garantidos, mais concretamente que o anonimato e a confidencialidade possam ser absolutamente defendidos.

Tipicamente existem dois tipos de riscos que potencialmente põem em causa esses princípios: a *desocultação da identidade*, ou a *desocultação do atributo* (DUPRIEZ; BOYKO, 2010). Presentemente estão disponíveis diversos métodos para prevenir o risco de identificação de uma unidade de registo que conste na base de microdados (DALE, 2001) e estão já identificados os principais cenários de divulgação dos dados onde existe maior risco para a manutenção da confidencialidade e anonimato da informação. Reconhecer, na situação de uso, qual o cenário de risco que está em causa é o primeiro passo para favorecer a opção por uma modalidade de anonimização dos dados. Os procedimentos adotados para este efeito não são independentes das características dos microdados em causa – os mesmos tanto podem se relacionar à natureza dos dados (censitários ou de amostragem), como às características do tipo de unidades de registo (empresas, edifícios, agregados familiares, indivíduos ou outras entidades). Quando estão em causa universos de grande dimensão a supressão das variáveis de identificação das unidades de registo constitui com frequência, por si só, um procedimento efetivo de garantia do anonimato. Mas mesmo este poderá não ser suficiente, tendo que se ativar outros procedimentos, sobretudo, quando o número de casos para certas categorias ou combinações de categorias é muito reduzido.

O quadro 2 sintetiza os principais modelos de anonimização utilizados para a disponibilização de microdados para a investigação científica.

## Quadro 2 - Modelos de anonimização de bases de microdados

Modalidade de anonimização	Descrição
A construção de <i>safe variables</i> : agregação de categorias	Consiste na seleção de variáveis contidas nas bases de dados originais (que poderiam conduzir à identificação de unidades estatísticas), as quais são sujeitas a um processo de anonimização por agregação de categorias. Este método pode ser utilizado quando não é possível assegurar a confidencialidade dos dados originais, por exemplo, devido ao nível de detalhe de uma determinada variável (unidade geográfica, profissional, etária, etc.).
A construção de <i>safe settings</i> : agregação de casos	Uma alternativa ao procedimento anterior é agregar, não as categorias de variáveis, mas os casos em subconjuntos, atribuindo a cada caso o valor de caracterização do subconjunto.
A construção de subamostras das bases de dados originais	A construção de subamostras dos dados originais, como forma de facilitar o acesso aos microdados em formato digital, ficando assegurado o anonimato.
A disponibilização de microdados em centros autorizados	Um outro modelo alternativo conjuga o <i>safe setting</i> com a disponibilização das bases de dados em centros autorizados onde os pesquisadores permanecem durante um determinado período para desenvolvimento de análises específicas. O <i>safe setting</i> assume aqui a forma de um cálculo laboratorial seguro, onde é rigorosamente impedida a alteração dos registos originais. Este método faculta a oportunidade de tratamento de dados de todos os casos coletados e não apenas de uma subamostra, como acontecia no caso anterior.
O acesso indireto aos dados	Outro processo que não envolve microdados, mas que em muitos casos resolve as necessidades dos utilizadores, consiste no acesso indireto aos dados através da elaboração de pedidos específicos, dirigidos à instituição produtora/gestora das bases de dados, de acordo com as análises que se pretendem produzir. Este sistema permite solicitar frequências ou cruzamentos entre indicadores.
O acesso através da Internet	Uma situação intermédia entre a referida anteriormente e os vários modelos de acesso direto aos microdados, consiste na disponibilização de bases de dados standardizadas na Internet. Este modelo possibilita a realização de apuramentos de frequências, seleção de categorias de variáveis e/ou segmentos da população, bem como, pedidos de cruzamentos com duas ou mais variáveis.

**Fonte:** Mauritti et al. (2002).

As soluções para dar resposta ao imperativo de anonimato no acesso a bases de microdados (ver quadro 1) são pois variadas. Um dos modelos mais utilizados diz respeito à construção de *safe variables* (agregação de categorias). Na seleção das variáveis a agregar tem-se em conta as perspetivas de análise e os temas específicos da pesquisa. Em certas circunstâncias pode ser vantajoso anonimizar com base numa agregação de âmbito geográfico, permitindo maiores especificações de variáveis educacionais ou profissionais, por exemplo, noutras, onde os indicadores de

caracterização espacial correspondam a vetores-chave de pesquisa, o procedimento poderá ir em sentido inverso, passando pela agregação de outros indicadores, nomeadamente níveis etários, grupos profissionais, etc.

A construção de *safe settings* (agregação de casos) é uma solução alternativa baseada no acesso em linha aos microdados. Neste modelo, os pesquisadores através de uma *password*, acedem, também, a uma subamostra representativa da base de sondagem. As variáveis dos vários inquéritos são previamente harmonizadas para permitir a comparabilidade (ANDROVITSANEAS, 2000; DALE, 2001; DUPRIEZ; BOYKO, 2010). Esta modalidade, face à anterior, apresenta algumas desvantagens – designadamente, o processo de agregação ser mais complexo e introduzir maior incerteza na análise. Para além disso o nível de interação com os dados é limitado, criando maiores obstruções à construção e teste de novas variáveis.

Um outro modelo diz respeito à *construção de subamostras das bases de dados originais* (ver quadro 1) e tem sido utilizado em países como o Reino Unido e aplicado sobretudo a *estatísticas censitárias* (DALE, 2001). Parte-se do princípio de que, com a construção destas subamostras, fica assegurada a impossibilidade de identificação dos dados. A informação assim disponibilizada é acessível tanto para fins de pesquisa como para o ensino.

A *disponibilização de microdados em centros autorizados* é um processo de preservação do anonimato, em geral, muito dispendioso, implicando a existência de uma equipe permanente que dê apoio e supervisione os pesquisadores. Alguns dos custos e inconvenientes recaem também sobre os pesquisadores, que têm de se deslocar para trabalharem com todos os dados, implicando, em certas situações, permanecer nesses centros até concluírem a pesquisa.

Por fim, os outros processos não dizem respeito, exatamente, à utilização de microdados, com todas as suas potencialidades. Por um lado, o acesso indireto aos dados, não permite, por exemplo, a criação de novas variáveis ou a elaboração de testes estatísticos sofisticados, designadamente de estatística multivariada, nem a experimentação indireta de sucessivos tratamentos alternativos, o que o torna menos atrativo. Por outro, a possibilidade de *acesso através da Internet*, para além de facultar a experimentação das bases de dados, tem a vantagem de integrar, na informação disponibilizada, um conjunto de variáveis derivadas (como a condição perante o trabalho, diversas taxas de emprego, etc.), cuja construção implica o domínio de algoritmos específicos. A nível nacional, esta poderá ser a orientação dominante na disponibilização de informação estatística por parte das instituições produtoras de microdados.

No acesso e uso de microdados estatísticos estão também implicadas as condições de disponibilização deste tipo de informação. Uma das questões essenciais diz respeito à

metainformação referente a cada projeto (FONSECA; VALENTE; LIMA, 1995, MAURITTI et al., 2002, DUPRIEZ; BOYKO, 2010). Esta refere a informação conceitual, metodológica e processual das bases de dados e às operações estatísticas que a ela conduziram.

Mas para que serve ao certo a metainformação? De acordo com Dupriez e Boyko (2010) esta tem as seguintes funções: a) ajudar os pesquisadores a *compreender* o que é que os dados medem e como são criados; b) *avaliar* a qualidade dos dados; e c) desenvolver *ferramentas de descoberta de dados*.

Para a utilização de bases de microdados é indispensável informação específica sobre os seguintes aspetos: a *metodologia do inquérito* (método de coleta, desenho do questionário, tipo e características da amostra, unidades estatísticas, tratamento de não-respostas, nomenclaturas); as *metodologias de estimação* (cálculo dos ponderadores e sua utilização); as *variáveis e seus conceitos* (descritivos das variáveis, incluindo as variáveis derivadas, tabelas de codificação e classificação, e condições de resposta ou filtros de seleção de casos); a *informação técnica* relativa às bases de dados (desenho de registo, estrutura das bases de dados, *software* utilizado e lista completa dos seus ficheiros); os procedimentos de *anonimização* (ver quadro 2); os critérios de *interpretação e divulgação* (metodologia de cálculo dos erros de amostragem associados aos resultados e utilização formatos de difusão) (MAURITTI et al., 2002; DUPRIEZ; BOYKO, 2010).

Outra preocupação na disponibilização dos dados refere-se à garantia da sua qualidade e à validação da informação produzida. É neste sentido, que se reconhece como fundamental o conhecimento de metodologias e procedimentos técnicos de coleta e tratamento por parte daqueles que utilizam a informação acedida: os *métodos de amostragem* e possibilidades de representatividade das amostras construídas – pressupõe que se tenha informação, sobre a população selecionada, a base de sondagem, a dimensão e métodos de seleção da amostra e critérios de estratificação que lhe estão associados, entre outros (BRYMAN, 2012; STURGIS, 2008) as formas de *cálculo dos ponderadores* – sendo muito relevante saber de que tipo de ponderador se dispõe e para que objetivos, nomeadamente se serve para corrigir a proporcionalidade no cálculo das relações entre indicadores e de que forma os valores ponderados são lidos em termos relativos ou em médias, ou ainda, para estimar os valores absolutos do universo; e os procedimentos de *validação sistemática de coleta, tratamento e análise dos dados* – onde os mais complexos e sensíveis são a eventual transposição da base de microdados de uma aplicação informática para outra (o que muitas vezes requer apoio técnico), a recepção dos dados e, como se referiu, a familiarização com a metainformação (MAURITTI et al., 2002).

Outros procedimentos são auxiliares de grande valor enquanto garantias de qualidade e validação dos dados. A comparação entre as questões (e modalidades de

resposta) que constam no instrumento de coleta de informação (questionários), e as variáveis presentes nas bases de dados disponibilizadas (e respectivas categorias) é um deles. Outro é o utilizador ter conhecimento acerca das mudanças ou quebras de séries dos inquéritos, que decorrem muitas vezes de aperfeiçoamentos na cobertura do inquérito em causa, rotações da amostra, correções sazonais e metodológicas (incluindo algumas alterações ao próprio questionário), etc. (ALLUM; ARBER, 2008; FONSECA; VALENTE; LIMA, 1995).

É também importante que se determinem alguns compromissos para a divulgação de informação e análise estatística por parte dos utilizadores, no sentido de garantir rigor tanto na informação como na sua referenciação. Este aspeto levanta dois tipos de problemas: um de *natureza ética* e um outro assente em procedimentos *técnicos*.

O primeiro, de ordem *ética*, refere-se às garantias de anonimização que o pesquisador deve dar quando difunde a informação (já aqui amplamente discutidas). A produção de conhecimento e sua divulgação estão também associadas à garantia de direitos e liberdades, nomeadamente na defesa da privacidade e anonimato individuais. O segundo, eminentemente *técnico*, relaciona-se com preocupações ligadas à qualidade dos dados, tais como o controle da significância, a construção de novas variáveis e a especificação, na divulgação destas, de designações distintas para não se confundirem com as variáveis oficiais.

As exigências de utilização de microdados alargam-se também às condições de difusão de análises apoiadas em estatísticas. Os requisitos deste tipo de prática dizem respeito, essencialmente, à necessidade de se garantir a qualidade e validade da informação (EUROSTAT, 2011). São elas: as características do processo de produção estatística, dos produtos estatísticos e de como é que os utilizadores percebem a qualidade desses produtos.

## HARMONIZAÇÃO E COMPARABILIDADE DOS DADOS

A comparação entre países é outra questão colocada quando se pondera a utilização e difusão de informação estatística. Tal aspeto, muito presente nas discussões da comunidade produtora, é concretizado, em grande parte, através da estandarização das metodologias de coleta de informação e/ou de variáveis entre os vários países, tendo necessariamente repercussões no uso de microdados.

Existe uma dupla vertente quando se considera a harmonização da informação estatística para efeitos de utilização de microdados. Uma centrada, essencialmente, na coleta e compilação dos dados (estratégia *input*), tem como vantagem permitir “controlar, de forma mais ponderada, constrangimentos relativos às especificidades de jurisdição e legislação nacionais face à regulamentação da produção e uso estatísticos,

bem como ao entendimento e cultura de informação que existe em cada país e Estado” (MAURITTI et al., 2002, p. 61). Outra refere-se ao tratamento de dados (estratégia *output*), nomeadamente quando se está a trabalhar no âmbito de estatísticas internacionais (COSTA, 2012; MARTINS, 2012), onde está em causa a uniformização de conceitos, exigindo aos seus utilizadores a interpretação dessa informação de forma contextualizada na diversidade social e cultural, das estruturas e dinâmicas sociais próprias de cada país.

A harmonização é a conjugação de estatísticas de fontes diversificadas, com o objetivo de tornar os conceitos e os métodos que as suportam partilháveis entre si. Este procedimento permite que a informação possa ser usada numa perspetiva longitudinal e, também, que seja possível construir análises comparativas. A utilização, por parte de pesquisadores, de bases de microdados e respetiva metainformação permite manter a perspetiva da comparabilidade, desde que se assegure que aqueles dominam as agregações implícitas às variáveis uniformizadas e controlam a qualidade imposta pela estandarização das variáveis. Ao mesmo tempo é possível enriquecer a abordagem comparativa com aprofundamentos de tipo regional ou nacional e, complementarmente, proceder-se à identificação de perfis construídos através de análises multivariadas, não prescindindo da qualidade dos conceitos comparados (MARTINS, 2012).

A vantagem de exercícios comparativos reside na forma como estes se constituem em instrumentos de decisões económicas e políticas e para um tipo de pesquisa que se pretenda situar numa escala europeia ou mundial (COSTA, 2012; MARTINS, 2012). O acesso aos microdados pode permitir, na comparação, enriquecer as análises, dotando-as de especificidades e aprofundamentos de âmbito nacional, testando os limites e a adequação dos conceitos harmonizados.

Os processos de harmonização implicam, muitas vezes, mudanças e interrupções dos inquéritos, o que cria constrangimentos a leituras longitudinais. Quando se utilizam bases de microdados esses aspectos também são sentidos, limitando as comparações entre as várias séries. Contudo a necessidade de aperfeiçoamento (com exigências na mudança metodológica dos inquéritos e dos conceitos) não pode ser travada, mesmo com alguns prejuízos nesse tipo de análises, e sabendo-se que parte da validação dessa informação depende da consistência da leitura de apuramentos numa perspetiva longitudinal.

A possibilidade de acesso a microdados, como a experimentada em Portugal, desde 1999, parece coincidente com outras vantagens postas à disposição, algumas *online*, nomeadamente as que se referem a informação comparativa entre países (como as fornecidas pelo EUROSTAT, OCDE ou UNESCO), e apresentada em bases de dados de indicadores, organizados em grandes temas, para cada país e cujos primeiros anos disponíveis remetem, precisamente, para o final dos anos 1990.

## UMA EXPERIÊNCIA DE 10 ANOS NO USO DE MICRODADOS

Uma política de acesso a microdados implica um entendimento estratégico do ponto de vista científico, um custo associado e, como vimos, uma capacidade técnica exigente para a sua implementação. O Protocolo estabelecido em Portugal para acesso a dados estatísticos para fins de pesquisa, diz respeito à formalização desse ensejo político e, até certo ponto, apresenta-se como resposta a uma reivindicação crescente por parte dos pesquisadores.

Passados mais de 10 anos sobre a abertura dessa possibilidade, importa analisar quem são os principais utilizadores, os respetivos enquadramentos institucionais e os temas e bases de dados predominantes dos pedidos.

O quadro 4 traça a dinâmica durante o período em referência (cerca de 12 anos), da qual se pode retirar algumas ilações. Em primeiro lugar, observe-se que o número de pedidos aumentou durante todo este período, embora tal não se verifique de forma linear. Nos últimos três anos completos, de 2010 a 2012, os pedidos situaram-se, em cada ano, num intervalo entre os 40 a 50 pedidos de bases de dados. Valores muito diferentes das solicitações iniciais – no primeiro triênio de vigência do Protocolo esses valores situam-se entre os 25 e 30 pedidos. Alguma variação nesta tendência pode ser explicada pelos períodos e ciclos de operação estatística que estão associados aos projetos do INE, recaindo os pedidos sobre dados atualizados no quadro de uma cadência própria na produção e divulgação dos dados de cada estudo. Estas flutuações associadas à divulgação de novos dados assinalam, em certo ponto, a relevância que os pesquisadores conferem à atualidade das estatísticas requeridas

No entanto, não é suficiente verificar o número total de pedidos para dar conta de algumas das tendências que marcam a sua composição. Um dos aspetos salientes é que no período em apreço os pedidos em equipe e os realizados de forma individualizada têm uma proporção equiparada. Ainda assim, a partir de evoluções destes dois perfis, é possível registar outras características do acesso aos microdados. Um aspeto interessante é o aumento exponencial de solicitações que se realizaram através de equipas, sobretudo a partir de 2007. Tal é sintomático de um movimento de grande institucionalização e profissionalização da ciência, e das ciências sociais em particular, nos últimos anos em Portugal (MACHADO, 2009). A sua configuração primordial tem assentado, sobretudo, na organização do trabalho científico em equipas. Por outro lado, a tendência, ainda que ligeira, de decréscimo de solicitações ao abrigo deste protocolo através de pesquisadores “isolados”, é um fenómeno relativamente recente e que ganha maior visibilidade a partir de 2009, contribuindo para a explicação anterior. É aliás nesta convergência analítica que se verifica um aumento de pedidos no quadro de projetos de pesquisa. Aspeto que adensa a proposição de transformação dos modos de fazer ciência, não só nas suas exigências técnicas e de complexificação do conhecimento produzido,

mas também do ponto de vista da sua institucionalização e da organização do trabalho científico.

### Quadro 3 - Dinâmicas de pedidos de bases de microdados em Portugal (2000-2012)

Pedidos no âmbito do Protocolo INE/MEC	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	Totais	%
<i>Nº de Pedidos</i>	26	30	27	33	41	32	25	28	39	36	48	41	47	453	100,0
Em equipa	3	11	6	13	16	8	8	17	17	27	34	27	36	223	49,2
Individuais	23	19	21	20	25	24	17	11	22	9	14	14	11	230	50,8
<i>Nº de Pedidos por tipologia</i>															
Projeto de investigação	14	12	9	15	17	14	13	21	20	20	33	31	28	247	54,5
Doutorado	9	16	14	14	16	8	11	6	14	8	10	3	9	138	30,5
Mestrado	3	2	4	4	8	10	1	1	5	8	5	7	10	68	15,0
<i>Nº total de Pesquisadores</i>	34	49	32	67	76	46	35	62	92	87	99	97	118	894	—

**Fonte:** INE 2013 (dados não publicados; cedidos à equipe de pesquisadores pelo Instituto Nacional de Estatística).

Os pedidos realizados para o desenvolvimento de doutorado parecem também ter conhecido algum recuo, sobretudo a partir de 2009 (quadro 3). Tal pode não significar uma diminuição da incorporação de informação de microdados no quadro da pesquisa fundamental desenvolvida neste tipo de graduação, mas antes que o seu desenvolvimento se enquadra em projetos de maior dimensão, refletindo, em certos casos, a dinâmica de pesquisa e formação avançada em centros de pesquisa. Os pedidos para o apoio à realização de mestrados não têm uma tendência muito nítida durante este período e ocupam uma parte mais reduzida do universo.

Do ponto de vista dos temas sobre os quais recaem mais os pedidos de microdados, a *demografia* e às *condições de vida das famílias* são os mais frequentes (ver quadro 4). As bases de dados mais requeridas são o Recenseamentos da População e Habitação (Censos) (23%), Inquérito aos Orçamentos Familiares (15%), Inquérito Anual às Empresas (IEH/SCIE)/Ficheiro de Unidades Estatísticas (FUE) (14%), Inquérito ao Emprego (14%). Com exceção do IEH/SCIE e dos Censos, cuja disponibilização de dados se faz através de um formato próximo de apuramentos especificados ou de subamostras agregadas (como subseções estatísticas), a informação relativa aos outros inquéritos é efetuada através de bases de microdados anonimizadas.

**Quadro 4 - Dinâmicas dos pedidos de bases de microdados estatísticos segundo os temas e os inquéritos requisitados**

Temas	Cód	Bases de Micro Dados e Apuramentos	2000/3	2004/7	2008	2009	2010	2011	2012	Total	%
1. Condições de vida das Famílias	1.1	Inquérito aos Orçamentos Familiares (IOF/IDEF)	26	44	8	4	9	4	7	102	15,2
	1.2	Indicadores de Conforto					1			1	0,1
	1.3	Painel Europeu Agregados Familiares (PEAF/ECHP)	15				5	1	1	22	3,3
	1.4	Inquérito às Condições de Vida e Rendimentos (ICOR/SILC)				3	6	4	6	19	2,8
	1.5	Inquérito à ocupação do tempo (IOT)				1	2			3	0,4
	1.6	Inquérito à Fecundidade					1	1	0	2	0,3
	1.7	Inquérito ao Património e Endividamento das Famílias (IPEF)				1	2	0	1	4	0,6
2. Empresas	2.1	Inquérito Anual às Empresas (IEH/SCIE) / Ficheiro de Unidades Estatísticas (FUE)	13	41		5	16	14	8	97	14,4
	2.3	Painel de Empresas					1	1	0	2	0,3
3. Comércio Internacional	3.1	Comércio Internacional				1	4	7	2	14	2,1
4. Construção de Edifício	4.1	Construção					2		0	2	0,3
5. Demografia	5.1/5.2	Óbitos	6	19			4	3	4	36	5,4
	5.3	Nados Vivos	5	20		3	4	3	4	39	5,8
	5.4	Casamentos					2	2	0	4	0,6
	5.5	Divórcios					1	2	0	3	0,4
	5.6	Recenseamentos da População e Habitação (Censos)	19	98	12	6	9	4	5	153	22,8
6. Mercado de trabalho	6.1	Inquérito ao Emprego (IE)	19	43	6	1	9	10	5	93	13,8
	6.2	Mód. Ad Hoc "Entrada dos jovens no mercado de trabalho -Inquérito ao Emprego" 2009						0	1	1	0,1
7. Agricultura	7.2	RGA/Inquérito Estrutura das Explorações Agrícolas					1	1	1	3	0,4
8. Indústria e Energia	8.1	Inquérito Anual à Produção Industrial (IAPI)				2	6	4	0	12	1,8
	8.1	Inquérito mensal à Industria						0	3	3	0,4
	8.2	Inquérito ao consumo de energia no setor doméstico					1	1	0	2	0,3
9. Sociedade de Informação	9.1	Inquérito à Utilização de Tecnologias da Informação e Comunicação (IUTIC)			6	2	5	5	3	21	3,1
10. Saúde	10.1	Inquérito Nacional de Saúde (INS)	6			2	6	3	3	20	3,0
	10.2	Inquérito aos Hospitais					1	1	3	5	0,7
11. Educação e Formação	11.1	Inquérito à Educação e Formação de Adultos (IEFA)					2	3	1	6	0,9
12. Cultura	12.1	Inquérito aos Museus						0	1	1	0,1
13. Ciências e Tecnologia	13.1	Inquérito Comunitário à Inovação (CIS)				1			1	2	0,3
Total			109	265	32	32	100	74	60	672	100

**Fonte:** INE 2013 (dados não publicados; cedidos à equipa de pesquisadores pelo Instituto Nacional de Estatística).

Observa-se ainda que os dois primeiros estudos mais solicitados se concentraram nos primeiros 7 anos de vigência do Protocolo e isso deve-se também às possibilidades de disponibilidade e atualidade dessas mesmas operações. Os últimos Censos tendo sido implementados no ano de 2011, só desde finais de 2012 é que estão disponíveis para solicitações dos pesquisadores e portanto ainda não foram objeto de muitos pedidos. Acrescente-se que na divulgação dos dados definitivos o INE colocou à disposição não apenas dos investigadores, mas do público em geral, uma enorme bateria de tabelas em formato *Excel*, com cruzamentos chave para análises investidas seja em caracterizações sociodemográficas, das estruturas familiares, recomposições socioprofissionais, qualificacionais, etc. A verdade é que mesmo tendo apenas por referência esses dados é já possível produzir análises de grande relevância sobre processos transversais de caracterização e mudança social.

Os dados provenientes dos vários projetos, disponíveis em bases de microdados permitem um conhecimento de tipo extensivo sobre um número muito alargado de indicadores que parecem cobrir o interesse dos mais variados temas trabalhados pelos cientistas sociais. Tal como se pode constatar no quadro 4, a partir de 2011, os inquéritos mais solicitados dizem respeito às empresas e ao mercado de trabalho. Tal resultado não deve ser alheio ao período de recessão e de dificuldades econômicas que Portugal atravessa.

A existência de uma agenda pública com efeitos políticos e na produção científica das ciências sociais, também é esperado que a mesma exerça pressão sobre a produção e difusão de informação ancorada em temas que adquirem em certos momentos maior centralidade pública.

## CONCLUSÃO

O acesso e utilização de microdados institucionais constituiu um importante marco no desenvolvimento das ciências sociais em Portugal. Esta possibilidade integra um movimento maior de acesso e partilha de instrumentos de pesquisa, bases dados e publicações referentes a resultados de pesquisa que está hoje em aberto e potenciado pelo recurso, entre outros, às tecnologias de informação e comunicação (CARDOSO; JACOBETTY; DUARTE, 2012).

A adesão a este tipo de possibilidade tem revelado um incremento consistente, sobretudo no que respeita ao desenvolvimento de estudos por equipas de trabalho com enquadramento em centros de pesquisa. No tratamento destes dados está a possibilidade de combinar novas propostas teóricas com a exploração de novas alternativas de recurso a informação de tipo quantitativo que dão conta de temas de forma atualizada e com possibilidades de leituras longitudinais.

Entre as potencialidades mais evidentes, a possibilidade de criação de novas variáveis e a oportunidade de realização de análises estatísticas multivariadas (incluindo a construção e aperfeiçoamento de tipologias e modelos de análise) são as mais importantes, dando conta de formatos de relação com esta informação baseados numa prática experimental e de uso continuado. Tal é verdade mesmo sabendo-se que nem todas as situações de pesquisa exigem ou tomam como indispensável a utilização de microdados.

A cobertura e as condições institucionais e organizacionais têm certamente influência sobre a eficiência e funcionamento técnico, bem como sobre a credibilidade das autoridades estatísticas que desenvolvem, produzem e divulgam estatísticas nacionais e internacionais. A multiplicação de análises sustentadas nas estatísticas oficiais vem, de fato, contribuir para reforçar a sua relevância e legitimidade enquanto principal fonte de informação. Na relação custos e benefícios, desde que salvaguardados todos os princípios éticos e técnicos que enquadram a coleta, produção, análise e difusão de dados, amplamente discutidos neste texto, todos beneficiam.

## REFERÊNCIAS

- ALLUM, Nick; ARBER, Sara. Secondary analysis of survey data. In: GILBERT, Nigel. (Ed.). *Researching social life*. London: Sage, 2008. p. 372-393.
- BRYMAN, Alan. *Social research methods*. Oxford: Oxford University Press, 2012.
- CARDOSO, Gustavo; JACOBETTY, Pedro; DUARTE, Alexandra. *Ciência aberta*. Lisboa: Mundos Sociais, 2012.
- ANDROVITSANEAS, Christos. New technologies in statistics and requirements for central institutions: from a user's perspective, innovation in provision and production of Statistics: The importance of new technologies. *Conferências do CEIES*, 9, 20-21 de Janeiro de 2000, Helsínquia. Disponível em <http://edz.bib.uni-mannheim.de/daten/edz-ma/eus/01/CEIES-9-1EN.doc>.
- COSTA, Antonio Firmino. Desigualdades globais. *Sociologia, Problemas e Práticas*, Lisboa, n. 68, p. 9-32, 2012.
- DALE, Angela. An academic perspective on the information society. In: CHALLENGES FOR COMMUNITY STATISTICS, 2001, Bruxelas. *Proceedings...* Bruxelas, 2001.
- DUPRIEZ, Olivier; BOYKO, Ernie. Dissemination of microdata files: principles, procedures and practices. *IHSN Working Paper*, Lisboa, n. 5, 2010.

EUROSTAT. *Código de conduta das estatísticas europeias*. Disponível em: <<http://ec.europa.eu/eurostat/quality>>. Acesso em: 2 jun. 2011.

FONSECA, Isabel Duarte; VALENTE, Isabel Lourenço; LIMA, José Pedro. A gestão da metainformação no sistema estatístico português. *Sociologia, problemas e práticas*, Lisboa, n.17, p. 153-168, 1995.

GABINETE DE PLANEJAMENTO, ESTRATÉGIA, AVALIAÇÃO E RELAÇÕES INTERNACIONAIS - GPEARI. *Produção científica Portuguesa: 1990-2010*. Disponível em:

<[http://www.dgeec.mec.pt/np4/%7B\\$clientServletPath%7D/?newsId=116&fileName=Indicadores\\_Bibl\\_2010.xls](http://www.dgeec.mec.pt/np4/%7B$clientServletPath%7D/?newsId=116&fileName=Indicadores_Bibl_2010.xls)>. Acesso em: 15 jun. 2011

MACHADO, Fernando Luís. Meio século de investigação sociológica em Portugal. *Sociologia*, Porto, n. 19, p. 283-343, 2009.

MARTINS, Susana da Cruz. *Escolas e estudantes da Europa*. Lisboa: Mundos Sociais, 2012.

MARTINS, Susana da Cruz; MAURITTI, Rosário; COSTA, Antonio Firmino. Padrões de vida na sociedade contemporânea. In: GUERREIRO, Maria das Dores Horta et al. (Org.). *Quotidiano e qualidade de vida, Portugal no contexto europeu*. Oeiras: Celta, 2007. v. III.

MAURITTI, Rosário. Padrões de vida na velhice. *Análise Social*, Lisboa, v. 39, n. 171, p. 339-361, 2004.

MAURITTI, Rosário; MARTINS, Susana da Cruz; COSTA, Antonio Firmino; ANTUNES, Ana Simões. Utilização de bases de microdados na investigação em ciências sociais. *Revista de Estatística do INE*, Lisboa, v. 2, p. 49-64, 2002.

PORTUGAL. Lei n. 22 de 13 de maio de 2008. Lei do Sistema Estatístico Nacional. *Diário da República*, n. 92, maio 2008. Disponível em: <[http://www.dgpj.mj.pt/sections/estatisticas-da-justica/legislacao/lei-n-6-89-de-15-de/downloadFile/file/L\\_22\\_2008.pdf?nocache=1210667806.0](http://www.dgpj.mj.pt/sections/estatisticas-da-justica/legislacao/lei-n-6-89-de-15-de/downloadFile/file/L_22_2008.pdf?nocache=1210667806.0)>. Acesso em: 5 maio 2013.

STURGIS, Patrick. Designing samples. In: GILBERT, Nigel (Ed.). *Researching social life*. London: Sage, 2008. p. 165-181.