A anonimização enquanto mecanismo de proteção de dados pessoais à luz da atual conjuntura legislativa europeia

Augusto Cesar Torbay*

Resumo: Consubstanciando uma exceção ao âmbito de aplicação material do RGPD, os dados anonimizados expressam uma solução de compromisso entre a tutela das pessoas singulares e a maleabilidade dos dados pessoais. Contudo, este Regulamento acaba por afastar-se de uma noção absoluta de anonimização, deixando-a aberta a uma ponderação de razoabilidade de meios. Partindo desta premissa, e observando que a concretização desse critério tem uma expressão direta no nível de proteção prestada aos respetivos titulares, procuraremos revisitar a questão da anonimização de dados pessoais no sentido de aferir da valência atual do respetivo regime, sem perder de vista os recentes contributos científicos que denunciam a sua incontornável fragilidade.

Palavras-chave: *Dados pessoais, Dados anonimizados, Anonimização de dados, Identificabilidade, RGPD.*

Abstract: Established as an exception to the material scope of the GDPR, anonymized data expresses a compromise between the protection of data subjects and the malleability of personal data. However, this Regulation departs from an absolute conception of anonymization, leaving it open to an assessment regarding the reasonable means likely

* Encarregado de Proteção de Dados junto da Autoridade Nacional de Segurança Rodoviária. Licenciado em Direito, Pós-graduado em Direito Comercial e Mestre em Direito Forense pela Faculdade de Direito da Universidade Católica Portuguesa de Lisboa, encontrando-se presentemente a frequentar o curso de Doutoramento na mesma instituição. Frequentou a 3.ª edição do Curso Breve sobre Proteção de Dados Pessoais organizado pela Associação da Faculdade de Direito da Universidade Nova de Lisboa (Jurisnova), bem como o curso de certificação de encarregados de proteção de dados promovido pelo European Centre on Privacy and Cybersecurity da Universidade de Maastricht. Iniciou a sua produção científica no campo da arbitragem internacional, encontrando-se, presentemente, a investigar e atuar na área da Tecnologia de Informação, Privacidade e Proteção de dados.

to be used to identify a natural person. Based on this premise, and observing that the execution of this criterion has a direct implication in the level of protection granted to data subjects, we intend to revisit the issue of anonymization of personal data in order to assess the current validity of the respective regime, without losing sight of the recent scientific contributions that denounce its unavoidable fragility.

***Keywords**: Personal data, Anonymized data, Data anonymization, Identifiability, GDPR.*

Considerações iniciais

Um dos principais traços característicos da ciência do Direito é a natureza dialética dos impulsos que promovem o seu desenvolvimento. Como que estabelecendo um diálogo com a realidade efetiva que procura regular, o Direito desenvolve-se no intento de acompanhar a complexidade dinâmica da interação do ser humano em comunidade. Fadada ao contínuo desfasamento e desatualização, esta correlação catalisa o progresso do Direito e promove o desenvolvimento de soluções justas. No entanto, e muito embora a evolução das interações humanas tenham seguido um ritmo que permitiu ao Direito ir acompanhando o compasso das realidades que visava conformar, o advento de um contexto social de cariz eminentemente tecnológico e de pendor tendencialmente global, parece antever que o concreto teor destas relações observará uma transmutação tal que dificilmente poderá ser objeto do mesmo nível de acompanhamento legal do que outrora se verificava.

Uma vez transposta para o contexto das operações enquadradas no universo digital, a tarefa da racionalização do pensamento jurídico vê-se impreterivelmente complexificada pelo carácter dinâmico e volátil das relações jurídicas que se formam no seu seio. O desenvolvimento contínuo de novos recursos e mecanismos de interação determina um panorama caótico, colocando o "legislador no intento constante de acompanhar o passo do desenvolvimento tecnológico"[1] na sua demanda por soluções

[1] BUTTARELLI, Giovanni. "Speech on "All we need is L....Privacy by design and by default", *Conferência RightsCon*, 2017, p. 4, disponível em: https://edps.europa.eu/data-protection/our-work/our-work-by-type/speeches-articles_en, acedido a: 27.02.2019.

que não sejam meramente justas, mas também flexíveis e ambivalentes, para assim lograr assegurar uma proteção eficaz do indivíduo.

Do nosso ponto de vista, encontramos uma expressão concreta desta dinâmica no enquadramento legal concedido pela legislação europeia ao regime da anonimização de dados pessoais. Transcorrido mais de um ano de aplicabilidade do RGPD[2], podemos afirmar que já se encontra enraizado na nossa consciência coletiva que, com este Regulamento, a UE procurou encetar um verdadeiro esforço de *empowerment* do titular dos dados.

Não deixa de ser igualmente evidente, porém, que o legislador europeu manteve presente que a proteção de dados não se subsume, como tal, a um valor absoluto, e que a realidade do presente enquadramento jurídico clama por um equilíbrio entre a proteção dos direitos fundamentais dos cidadãos e a promoção da liberdade dos atores comerciais que operam e promovem o ecossistema digital. Como que expressando uma solução de compromisso, e no sentido de permitir uma melhor maleabilidade de informações, o RGPD vem consagrar um regime de excecionalidade aos dados anonimizados, coartando-os da sua tutela e mantendo-os aquém dos direitos e obrigações que reconhece ao tratamento de dados pessoais.

Conforme procuraremos demonstrar no presente estudo, consideramos que é precisamente aqui que reside o desencontro entre o progresso legislativo e o substrato tecnológico que, em última análise, ameaça a coerência da tutela europeia consagrada à proteção de dados pessoais. O carácter excecional concedido às informações produzidas por um processo de anonimização decorre do prejuízo da identificabilidade destes dados, porém, somos crescentemente encarados com o facto de que as novas tendências computacionais[3], bem como a crescente disponibilização de dados públicos, contribuem para que a produção de dados eficientemente anonimizados se revista de uma crescente complexidade.

---

[2] Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE.

[3] OHM, Paul. "Broken promises of privacy: Responding to the surprising failure of anonymization", *UCLA Law Review*, v. 57, 2010, p. 1701-1777, disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006, acedido a: 21.02.2017.

Muito embora, a nível conceptual, o processo de anonimização pudesse consubstanciar o perfeito equilíbrio entre a proteção do titular dos dados e a maleabilidade da informação, ao arrepio da confiança que o legislador europeu lhe reconheceu, torna-se cada vez mais evidente que existe uma crescente unicidade de vozes que se levanta relativamente ao declínio da valência da anonimização de dados pessoais e que denuncia uma tendência para a efetiva impossibilidade de produção de dados que, além de anónimos, logrem suster utilidade.

Nesta medida, pretendemos dedicar o presente estudo à observação do regime dos dados pessoais anonimizados, procurando ponderar a valência da conceptualização do processo de anonimização, bem como a sua efetividade na tutela dos direitos dos titulares dos dados. Aproveitaremos também o presente ensejo para, à luz dos contributos científicos evidenciados ao longo dos últimos anos, aferir se poderia ser efetivamente pertinente uma reponderação a nível da abordagem europeia em relação a esta matéria[4].

1. O regime jurídico dos dados pessoais anonimizados no âmbito do RGPD

A base fundamental da aplicação material da tutela europeia relativa à proteção de dados pessoais (a qual é determinada pela conjunção do disposto no n.º 1 do art.º 2 com a noção estatuída pelo n.º 1 do art.º 4 do RGPD), subsume-se ao simples facto de que as disposições do Regulamento apenas serão aplicáveis ao tratamento de informações "relativa[s] a uma pessoa singular identificada ou identificável", entendendo-se, esta última, como uma "pessoa singular que possa ser identificada, direta ou indiretamente".

A consequência lógica que importa retirar desta abordagem é que, se uma determinada informação não concernir a uma pessoa singular

---

[4] ZIBUSCHKA, Jan; KUROWSKI, Sebastian; ROßNAGEL, Heiko; SCHUNCK, Christian H; ZIMMERMANN, Christian. "Anonymization Is Dead – Long Live Privacy" in ROßNAGEL, Heiko (Ed.); WAGNER, Sven (Ed.); HÜHNLEIN, Detlef (Ed.). *Gesellschaft für Informatik*, 2019, p. 71-82, disponível em: https://dl.gi.de/bitstream/handle/20.500.12116/20995/proceedings-06.pdf, acedido a: 23.07.2019.

identificada ou identificável, esta informação deverá ser considerada anónima e, como tal, não será compreendida no escopo material do RGPD. Contudo, não são apenas as informações anónimas que cabem neste regime. Nos termos avançados pelo considerando 26 deste Regulamento, considerar-se-ão igualmente desenquadradas do seu regime todos os dados que, embora possam ter sido considerados como pessoais, tenham sido "tornados de tal modo anónimos que o seu titular não seja ou já não possa ser identificado". Na prática, isto se traduz no facto de que tanto os dados anónimos, como os anonimizados (ou seja, dados pessoais que foram submetidos a um processo de anonimização), não caberão no âmbito de aplicação material do RGPD.

Numa primeira análise, em termos estritamente conceptuais, esta determinação aparenta assentar as bases do processo de anonimização numa presunção de impossibilidade absoluta de reidentificação (consubstanciando um conceito de *guaranteed anonymisation*, conforme concetualizado pela doutrina[5]). Tal como sugere o parecer do Grupo de Trabalho sobre técnicas de anonimização[6], muito embora não exista uma norma prescritiva na legislação europeia que estatua a forma como este processo deve ser realizado, na sua abordagem encontra-se subjacente uma determinada notação de irreversibilidade[7].

Aliás, do nosso ponto de vista, e através de um exercício de interpretação sistemática, entendemos que o regime jurídico que é consagrado no quadro da legislação europeia de proteção de dados permite supor a propensão do legislador para uma conceção de anonimização "tão permanente quanto a eliminação"[8]. Nomeadamente, no âmbito do RGPD, esta correlação encontra expressão na própria consagração do princípio da limitação da conservação dos dados pessoais (*cfr.* alínea e) do n.º 1

---

[5] OHM, Paul. *op. cit.*, p. 7.

[6] Grupo de Trabalho, Parecer 05/2014 sobre técnicas de anonimização, 0829/14/PT, 10/03/2014, p. 7, disponível em: https://www.gpdp.gov.mo/uploadfile/2016/0831/20160831042518381.pdf, acedido a: 06.02.2019.

[7] Em certos domínios, a relação entre o processo de anonimização de dados pessoais e a exigência de irreversibilidade é expressa literalmente, como seja, nomeadamente, o caso da norma 29100:2011, que vem determinar que as informações sobre "pessoas identificáveis" devem ser "alteradas irreversivelmente" de modo a que as mesmas "já não possa[m] ser identificada direta ou indiretamente".

[8] Grupo de Trabalho, *op. cit.*, p. 6.

do art. 5.º). Ao contrário do que poderia ser empiricamente expectável, esta norma não determina a obrigatoriedade de eliminação dos dados pessoais após o "período necessário para as finalidades para as quais são tratados". O que efetivamente se estatui é a proibição da sua conservação numa "forma que permita a identificação dos titulares dos dados". Em concreto, o que se determina nesta disposição é que, de acordo com o princípio da limitação da conservação dos dados pessoais, após o termo do referido período, o responsável pelo tratamento deverá proceder ao apagamento dos dados ou, em alternativa, promover a sua anonimização. Salvo melhor entendimento, ao estatuir a anonimização como alternativa à eliminação, o legislador europeu demostra a expectativa de um processo de anonimização tão eficiente quanto a eliminação dos dados.

Mutatis mutandis, observamos esta mesma equiparação no regime consagrado ao tratamento de dados de tráfego no sector das comunicações eletrónicas. Conforme o n.º 1 do art.º 6 da Diretiva 2002/58/CE[9], estes dados "devem ser eliminados ou tornados anónimos" quando deixem de ser necessários para efeitos da transmissão da comunicação aos quais se referem. Por sinal, e nos termos da Proposta de Regulamento referente a esta matéria e que se prevê que venha a revogar a referida Diretiva 2002/58/CE[10], este mesmo entendimento será consagrado, não apenas quanto aos dados referentes ao conteúdo das comunicações eletrónicas, como também em relação aos respetivos metadados[11]. Por outro lado, importa salientar que a equiparação da anonimização à eliminação não é apenas transversal ao direito da União, verificando-se ainda a sua influência para lá das fronteiras da Europa. Nomeadamente, uma postura semelhante encontra-se vertida nos termos da alínea e) do n.º 4 do art.º 5.º da Convenção para a Proteção das Pessoas Relativamente ao Tratamento Automatizado de Dados de Caráter Pessoal do Conselho da Europa.

---

[9] Diretiva n.º 2002/58/CE do Parlamento Europeu e do Conselho de 12 de julho de 2002 relativa ao tratamento de dados pessoais e à proteção da privacidade no sector das comunicações eletrónicas.

[10] Proposta de Regulamento do Parlamento Europeu e do Conselho relativo ao respeito pela vida privada e à proteção dos dados pessoais nas comunicações eletrónicas e que revoga a Diretiva 2002/58/CE.

[11] *Cfr.* n.º 1 e 2 do art. 7.º da Proposta de Regulamento.

Com base nestas evidências, entendemos que seria possível supor que o carácter de irreversibilidade se encontraria no âmago da noção europeia de anonimização, como *conditio sine qua non* da sua efetividade[12]. Não obstante, através dos contributos dos respetivos considerandos, o RGPD vem afastar-se desta noção estrita de anonimização[13]. Procurando favorecer uma conceção mais maleável, o legislador europeu estatui uma solução de compromisso através de uma conceptualização fundamentada numa ponderação de razoabilidade e cujos vetores se encontram delimitados pelo próprio conceito de identificabilidade[14].

No contexto do mencionado considerando 26, o Regulamento vem determinar que, para que um dado possa ser considerado anonimizado, não será necessário que o respetivo titular "não seja ou já não possa ser identificado" em termos absolutos. Em concreto, para efeitos desta abordagem, bastará que, observando "todos os meios suscetíveis de ser razoavelmente utilizados (...) quer pelo responsável pelo tratamento quer por outra pessoa" o respetivo titular dos dados não possa ser direta ou indiretamente identificado. Ou seja, e conforme referido acima, a chave da ponderação centra-se na avaliação do que é "razoável" de ser utilizado para efeitos da identificação. O que verdadeiramente ocorre perante esta consagração é uma flexibilização do conceito, o qual, necessariamente, fica na dependência de um juízo de ponderação dos recursos disponíveis para a reversão do processo.

Enquanto direito fundamental, o direito à proteção de dados pessoais não é, em si, absoluto. Estando aberto à ponderação de juízos de

---

[12] Grupo de Trabalho, *op. cit.*, p. 7.

[13] Tal como refere o Advogado-Geral Maciej Szpunar nas conclusões apresentadas em 21 de março de 2019 no âmbito do processo C-673/17 do TJUE, muito embora, na ordem jurídica da União, os considerandos se encontrem desprovidos de natureza prescritiva e, como tal, não possuam valor jurídico independente, a verdade é que representam uma ferramenta obrigatória na interpretação das disposições de um ato jurídico da união (Conclusões do Advogado-Geral Maciej Szpunar de 21 de março de 2019, *Planet49 GmbH contra Bundesverband der Verbraucherzentralen und Verbraucherverbände – Verbraucherzentrale Bundesverband e.V.* C-673/17, ECLI:EU:C:2019:246, para. 71).

[14] PURTOVA, Nadezhda. "The law of everything. Broad concept of personal data and future of EU data protection law". *Law, Innovation and Technology*. v. 10, n.º 1, 2018, p. 40-81, disponível em: https://doi.org/10.1080/17579961.2018.1452176, acedido a: 17.03.2019.

proporcionalidade[15], compreende-se a intenção do legislador europeu em favorecer uma relação de compromisso entre a tutela deste direito e os interesses na maleabilidade dos dados. Contudo, a abertura deste conceito a uma lógica assente num critério de razoabilidade torna evidente que uma dimensão de risco será sempre uma realidade inerente ao próprio processo de anonimização[16], pelo que, permitimo-nos suspeitar que esta sujeição à indeterminação poderá representar uma fragilidade que não se coadune com o regime de excecionalidade que é consagrado aos dados anonimizados à luz do RGPD.

Esbatendo a fronteira que demarca a anonimização da pseudonimização[17] (a qual, no seu cerne, se caracteriza pela possibilidade de reidentificação através de um atributo único) cria-se um *tertium genus* localizado entre ambos os conceitos e que, no fundo, se carateriza pela nomenclatura do primeiro e pelo risco inerente do segundo.

No essencial da sua conceptualização, deveria ser esta dimensão de risco que demarcaria a anonimização da pseudonimização. Ao reduzir a possibilidade de "ligação entre um conjunto de dados e a identidade original de um titular", esta última, representa uma "medida de segurança útil, porém, não consubstancia um método de anonimização de dados pessoais"[18]. Conforme evidenciado pelas investigações realizadas em torno do já clássico caso de publicação de registos da Comissão de Táxis e Limusines de Nova Iorque em 2014[19], o recurso à pseudonimização

---

[15] HANSSON, Mats; LOCHMULLER, Hanns; RIESS, Olaf; SCHAEFER, Franz; ORTH, Michael; RUBINSTEIN, Yaffa; MOLSTER, Caron; DAWKINS, Hugh; TARUSCIO, Domenica; POSADA, Manuel; WOODS, Simon. "The risk of re-identification versus the need to identify individuals in rare disease research", *European Journal of Human Genetics.* 24, 2016, p. 1553-1558, disponível em: https://www.ncbi.nlm.nih.gov/pubmed/27222291, acedido a: 17.03.2019.

[16] Grupo de Trabalho, *op. cit.*, p. 7.

[17] A qual, nos termos do n.º 5 do art. 4.º do RGPD, se concretiza no tratamento de dados tendente a impossibilitar a sua atribuição a um titular de dados específico sem o auxilio de informações suplementares, as quais, deverão ser mantidas separadamente e sujeitas a medidas técnicas e organizativas que assegurem a permanência da pseudonimização dos dados.

[18] Grupo de Trabalho, *op. cit.*, p. 22.

[19] Em 2014 a comissão de táxis e limusines de Nova Iorque publicou uma base de dados com registos relativos a todas as viagens realizadas nessa cidade durante um ano. Com intuito de promover a proteção dos titulares dos dados envolvidos, procurou-se realizar a pseudonimização através da utilização de uma função criptográfica *hash* sobre determinados elementos, em particular, o número identificativo da viatura e o número da licença do respetivo

detém riscos próprios decorrentes da efetiva possibilidade de reidentificação[20]. Em consequência deste facto, e por consciência deste risco, o regime jurídico dos dados pseudonimizados não os afasta plenamente da tutela do RGPG.

Na nossa perspetiva, e avançando considerações que procuraremos explorar *infra*, consagrar um conceito de anonimização que se caracteriza, necessariamente, por uma dimensão omnipresente de risco poderá revelar-se particularmente problemática, principalmente se mantivermos presente que os dados anonimizados contornam a tutela consagrada pelo RGPD. Muito embora este Regulamento declare a sua intenção de prestar uma tutela real e efetiva aos titulares dos dados, consideramos que a consagração de uma noção frágil de anonimização poderá concorrer para o desenvolvimento de um ambiente de insegurança jurídica, como resultado da possível frustração das expectativas depositadas na viabilidade das garantias prestadas pelo processo de anonimização.

2. A viabilidade efetiva da conceção europeia de anonimização de dados pessoais

Assente na ideia de impossibilidade prática de alcançar dados absolutamente anonimizados, e procurando obviar um regime demasiado estrito, o legislador europeu concebeu uma solução fundamentada num entendimento relativo de anonimização. Caberá, no entanto, questionar se esta abordagem é efetivamente operacional ou se, pelo contrário, a

---

condutor. Como a função *hash* representa uma relação de muitos para um, a complexidade na sua reversão encontra-se na dimensão do universo de pesquisa na qual se enquadra. Quando o *input* da função é considerável, a reversão vê-se complexificada e, consequentemente, são promovidas as garantias de segurança. Neste caso, porém, o universo em questão, ou seja, o número de condutores de táxis, embora empiricamente substancial, não representou obstáculo definitivo à descoberta do algoritmo de pseudonimização, tendo-se logrado a reversão com o recurso a informações publicamente disponíveis.

[20] NALDI, Maurizio. "Anonymization Systems and Utility", conferência IPEN de 12 de junho de 2019, sobre o tema *"State of the art" in data protection by design – Current state and future trends*, Itália. Apresentação disponível em: https://edps.europa.eu/sites/edp/files/publication/12-06-19_maurizio-naldi_anonymization-systems-and-utility_en.pdf, acedida a: 27.07.2019.

indeterminação do critério ao qual se submete se reveste de uma complexidade tal que, na prática, acabe por desembocar igualmente numa condição impossível de atingir.

No sentido de nos permitir tecer considerações concretas quanto à efetiva viabilidade da noção europeia de anonimização de dados, e assumindo como premissa o critério da razoabilidade de meios consagrado pelo RGPD, no presente ponto, procuraremos analisar – de forma necessariamente sumária – aqueles que se nos evidenciam como os principais obstáculos e dificuldades à produção de dados eficientemente anonimizados.

2.1. A inexistência de um padrão determinado para um processo de anonimização eficiente

Ao colocar as bases do processo de anonimização num critério de razoabilidade de meios, o legislador vem determinar que a anonimização deverá considerar-se alcançada apenas quando o panorama do contexto concreto do tratamento e as circunstâncias específicas que influem sobre a identificabilidade permitam considerar que a reidentificação se tornou "razoavelmente impossível"[21]. Contudo, além da conceptualização deste critério, parece-nos indiscutível que não dispomos de um padrão viável que nos permita antever qual o procedimento apropriado para lograr esse nível de anonimização. Da mesma forma, não existe uma predeterminação que indique quais os recursos considerados necessários para a reversão ou que permita a construção de um modelo concreto de aferição da robustez de uma determinada estratégia de anonimização[22].

A verdade, porém, é que a inexistência de tal roteiro não se encontra inteiramente desprovida de sentido. Se atendermos ao facto de que esta área é um campo em pleno desenvolvimento, cujas fronteiras se encontram em constante reformulação em decorrência da contínua investigação de que é alvo[23], facilmente se compreende que uma concretização de um modelo uniforme de anonimização não se revelaria como uma

---

[21] Grupo de Trabalho, *op. cit.*, p. 9.

[22] *Ibid.*, p. 30.

[23] PURTOVA, Nadezhda, *op. cit.*, p. 78.

efetiva mais-valia pois, rapidamente, observar-se-ia a sua condenação à desadequação prática e consequente ineficácia. Principalmente por se fundamentar numa ponderação de razoabilidade, o conceito de anonimização consagrado no âmbito do RGPD torna-se uma realidade dinâmica que encontra os seus fundamentos operacionais na capacidade informática de processamento de informação e que, como tal, exige uma ponderação flexível e contínua, que não se coaduna com uma medida uniforme e estática de anonimização.

No entanto, o carácter flexível desta abordagem tem como consequência necessária a indeterminação do conceito de anonimização que, em determinados casos, se apresenta como um elemento incontornável. É precisamente neste sentido que vão proliferando na doutrina[24] casos de estudo que evidenciam a extrema dificuldade de criar uma base de dados com um nível de anonimização satisfatório[25].

Ao entregar o conceito de anonimização de dados pessoais a uma ponderação de razoabilidade de meios de identificação, a legislação europeia vem determinar uma dimensão de complexidade ao procedimento que poderá, em última análise, ser intransponível. Nos termos desenvolvidos pelo referido considerando, a avaliação da razoabilidade da reversão do processo deverá ser realizada numa perspetiva eminentemente global, atenta à uma universalidade de elementos dinâmicos e voláteis, que deverá ser mantida ao longo do tempo. Perante uma tal exigência, a inexistência de vetores de orientação concretos poderá entregar esta ponderação a uma dimensão necessariamente subjetiva, de validade efémera e, nessa medida, propensa à incerteza e ao risco.

2.2. A universalidade de elementos que influem na avaliação do critério da razoabilidade de meios

Concretizando aquele que é um dos poucos vetores de orientação que o Regulamento nos presta nesta matéria, o seu considerando 26 procura

---

[24] ROCHER, Luc; HENDRICKX, Julien M.; MONTJOYE, Yves-Alexandre de. "Estimating the success of re-identifications in incomplete datasets using generative models". *Nature Communications.* v. 10, n.º 3069, 2019, p. 2, disponível em: https://www.nature.com/articles/s41467-019-10933-3, acedido a: 30.07.2019.

[25] Grupo de Trabalho, *op. cit.*, p. 3.

evidenciar que todo o processo de anonimização de dados pessoais estará pendente de uma avaliação eminentemente multidimensional. A premissa, como já sabemos, é a ponderação transversal que o responsável deverá encetar e que, necessariamente, terá de compreender a totalidade dos meios razoavelmente utilizáveis para efeitos de identificação direta ou indireta. Contudo, na sua concretização, esta avaliação não se poderá limitar à perspetiva do próprio responsável pelo tratamento, uma vez que deverão ser trazidos à colação os recursos identificativos de eventuais terceiros que possam ter acesso legal aos dados em causa.

Em termos concretos, o que se deve retirar desta construção é que o legislador comunitário procurou consagrar uma abordagem integral ao processo de anonimização, estatuindo como que um "dever de meios" (implementação dos meios considerados razoáveis para garantir a robustez da anonimização) mais do que de um "dever de resultados"[26] (produção de dados absolutamente anónimos). Através de um exercício de ponderação global, que não considere apenas elementos intrínsecos à informação em causa (como seja, nomeadamente, a própria natureza dos dados a tratar), mas igualmente fatores eminentemente extrínsecos (como o enquadramento contextual do tratamento), o legislador da União pretende que o responsável possa garantir que, embora não seja impossível, a identificação dos dados implicaria um esforço desrazoável.

Do nosso ponto de vista, a transversalidade exigida à ponderação dirige-nos à conclusão de que o sucesso da anonimização dependerá sempre da relação entre as particularidades dos próprios dados e os elementos contextuais que enquadram o tratamento[27]. Nesta linha de raciocínio, pretendemos dedicar o presente título à observação destes elementos, no sentido de aferir da sua influência no processo de anonimização.

---

[26] ELLIOT, Mark; MACKEY, Elaine; O'HARA, Kieron; TUDOR, Caroline. *The Anonymisation Decision-Making Framework*, University of Manchester, GB. UKAN, 2016, p. 18.

[27] *Ibid.*, p. 52.

2.2.1. A influência de elementos de carácter extrínseco: A consideração de fatores contextuais do processo de anonimização

Revestindo-se a anonimização de um cariz complexo e interrelacionado, torna-se evidente que uma avaliação de elementos contextuais (como sejam, nomeadamente, os agentes envolvidos, o tipo de procedimentos implementados ou as infraestruturas disponíveis) seja um elemento essencial à sua ponderação. Aliás, o carácter evolutivo e dinâmico da ciência da computação, bem como a sua capacidade de processamento de informação e, consequentemente, a complexificação dos dados objeto de tratamento, exigem que o respetivo responsável se retenha especialmente na avaliação do contexto tecnológico que compagina o conjunto dos meios suscetíveis de serem razoavelmente utilizados para efeitos de identificação. Como consequência direta desta relação, é hoje evidenciado pela doutrina que o desenvolvimento tecnológico enforma a noção de privacidade e confidencialidade[28]. Mais, em virtude da contínua investigação levada a cabo neste campo, assiste-se a um constante redefinir das fronteiras do que é razoavelmente expectável de ser realizado, observando-se uma continua potencialização de recursos, diminuição de custos e aumento de especialização[29].

Chegamos a encontrar opiniões que denunciam que, na maior parte das bases de dados confiadas como anonimizadas, a identificação não requer mais do que conhecimentos de estatística básica e programação[30]. Desta forma, facilmente se antevê a possibilidade de que determinadas avaliações (realizadas, naturalmente, no momento da anonimização), possam vir a ser postas em causa e, consequentemente, prejudicar a

---

[28] ZIMMERMANN, Christian; CABINAKOVA, Johana; "A Conceptualization of Accountability as a Privacy Principle" in ABRAMOWICZ W. (eds) *Business Information Systems Workshops. Lecture Notes in Business Information Processing*, v. 228, 2015, p. 261–272, disponível em: https://www.semanticscholar.org/paper/A-Conceptualization-of-Accountability-as-a-Privacy-Zimmermann-Cabinakova, acedido a: 22.08.2019.

[29] Grupo de Trabalho, *op. cit.*, p. 9.

[30] NARAYANAN, Arvind; FELTEN, Edward W. *No silver bullet: De-identification still doesn't work*, 2014, p. 6, disponível em: https://iapp.org/resources/article/no-silver-bullet-de-identification-still-doesnt-work, acedido a: 27.07.2018.

proteção dos respetivos titulares[31]. Apenas considerando o contexto tecnológico no qual se enquadra o tratamento é possível tentar alcançar uma relação eficiente entre os esforços necessários à anonimização de dados pessoais e os recursos exigidos para a sua reversão (como o custo, o saber-fazer ou o tempo)[32].

Uma falácia comum, que assenta na desconsideração do carácter evolutivo do enquadramento tecnológico, é julgar que a anonimização é um estado determinado em que um particular conjunto de dados se encontra[33]. Muito pelo contrário, a verdade é que o ritmo acentuado do desenvolvimento "da capacidade computacional e das ferramentas disponíveis"[34] demanda uma aproximação dinâmica a esta questão e, principalmente, o reconhecimento de que a anonimização possui uma propriedade eminentemente volátil, que não deve ser apenas atingida, como também mantida no tempo.

As considerações de cariz contextual, porém, não se devem limitar à avaliação do *status quo* dos recursos tecnológicos, mas também ao manancial de informação disponível que poderá influenciar a reidentificação da informação anonimizada. Tal como evidenciado por múltiplas investigações, encontra-se hoje perfeitamente demostrada a efetiva possibilidade de identificar informação anonimizada com recurso a bases de dados públicas[35].

Devemos manter presente que na avaliação da razoabilidade de meios disponíveis para a identificação não está apenas em causa uma abordagem relativa do conceito de identificabilidade. Como vimos, para efeitos de reidentificação direta ou indireta, a ponderação de razoabilidade não deverá ser realizada apenas da perspetiva do responsável pelo tratamento[36].

---

[31] PORTER, C. Christine. "De-identified data and third party data mining: The risk of reidentification of personal information", *WashingtonJournal of Law, Technology & Arts,* 5, 2008, p. 3, disponível em: http://www.lctjournal.washington.edu/Vol5/a03Porter.html, acedido a: 27.07.2019.

[32] Grupo de Trabalho, *op. cit.*, p. 10.

[33] ELLIOT, Mark. *et al*, *op. cit.*, p. 1.

[34] Grupo de Trabalho, *op. cit.*, p. 7.

[35] HANSSON, Mats. *et al*, *op. cit.*, p. 1554.

[36] *Cfr. o* considerando 26 do RGPD.

Aprofundando estas considerações, o TJUE[37] veio consagrar que o carácter de identificabilidade não dependerá apenas dos meios próprios de uma determinada pessoa, mas deverá ser observada também a suscetibilidade de identificação decorrente da combinação com outros recursos aos quais se possa legitimamente vir a ter acesso[38].

A conclusão necessária desta determinação é que a ponderação de razoabilidade de meios não se subsume apenas a uma avaliação de carácter endógeno, que se concretize na consideração da capacidade identificativa do responsável pelo tratamento em relação aos próprios meios. Uma vez que a anonimização decairá, não apenas com a recuperação plena e precisa dos dados do titular, mas também com a mera identificação, ligação ou inferência decorrente de outras fontes, tanto públicas como privadas, a robustez da anonimização estará sempre pendente de uma ponderação de carácter exógeno.

A questão é que esta aceção do conceito de identificabilidade agrava a complexidade do processo de anonimização devido à "crescente disponibilidade pública de outros conjuntos de dados"[39]. Um dos problemas clássicos na elaboração de bases de dados anónimas é a incapacidade de considerar corretamente a influência dos dados publicamente disponíveis, que podem influir decisivamente na identificação dos dados anonimizados, resultando em casos de processos de anonimização incompletos que "comportam consequências adversas, e por vezes irreparáveis, para os titulares dos dados"[40].

Desde a clássica demonstração realizada por Latanya Sweeney em 1995[41] até casos mais recentes, como a identificação de políticos alemães

---

[37] Neste sentido, o Acórdão do TJUE de 19 de outubro de 2016, Patrick Breyer v Bundesrepublik Deutschland, C-582/14, EU:C:2016:779.

[38] Neste mesmo sentido, o Acórdão do TJUE de 20 de dezembro de 2017, *Peter Nowak v Data Protection Commissioner*, C-434/16, ECLI:EU:C:2017:582.

[39] Grupo de Trabalho, *op. cit.*, p. 9.

[40] *Ibid.*

[41] No contexto da sua investigação, e através da conjugação de registos médicos publicados pelo governo da Massachusetts com informação publicamente disponível, a autora logrou destacar os registos médicos do governador William Weld, utilizando elementos identificadores como o sexo, data de nascimento e código postal (os quais, igualmente, se encontravam publicamente disponíveis). Por outro lado, a própria confirmação da correta identificação do governador foi realizada mediante a contraposição dos dados obtidos com outras informações

com base em histórico de pesquisas online[42] ou a identificação de cidadãos australianos a partir da publicação de registos médicos[43], evidencia-se a existência de uma fragilidade inerente à anonimização que se encontra na possibilidade da sua conjugação com bases de dados publicamente disponíveis.

Para procurar fazer face a esta debilidade, o processo de anonimização nunca se poderá resumir à eliminação de identificadores diretos que impossibilitem a sua reversão por parte do próprio responsável. Estando assente que, em última análise, a eficácia[44] da anonimização está também dependente dos recursos identificativos de terceiro, o seu processo deverá passar, necessariamente, pela cuidada consideração, não apenas as bases de dados de carácter privado no domínio de terceiros, mas também das informações públicas que, eventualmente, possam ser imiscuídas com os dados anonimizados[45].

Torna-se, assim, evidente que existe uma universalidade de fontes de informação que deverão ser consideradas na ponderação da robustez de uma determinada anonimização. Com efeito, foi precisamente a incapacidade de considerar esta dimensão global que esteve na base do

---

públicas referentes ao mesmo (SWEENEY, Latanya. "K-anonymity: A model for protecting privacy". *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, v. 10, n.º 05, 2002, p. 557-570).

[42] Entre os exemplos discutidos na doutrina, um dos mais mediáticos é o caso da investigação levada a cabo pela jornalista Svea Eckert que, em 2016, logrou individualizar o histórico de pesquisa de políticos alemães (bem como de outras figuras públicas) com base em dados anonimizados referentes a históricos de pesquisa de 3 milhões de cidadãos alemães, aos quais teve acesso de forma gratuita.

[43] Uma equipa da universidade de Melbourne logrou realizar um exercício de identificação com base em registos médicos anónimos de 2.9 milhões de cidadãos australianos. Os referidos dados tinham sido publicados pelo departamento de saúde australiano no âmbito de uma política de transparência em agosto de 2016 sobre a premissa da sua anonimidade. Entre outros dados, o registo constava de informações referentes a diagnósticos, tratamentos e respetivos custos.

[44] Nos termos da esquematização realizada pelo Grupo de Trabalho, a eficácia de uma determinada estratégia de anonimização deverá ser aferida pela sua capacidade de impedir "que qualquer uma das partes identifique uma pessoa num conjunto de dados, relacione dois registos num conjunto de dados (ou entre dois conjuntos de dados separados) e deduza quaisquer informações desse conjunto de dados" (Grupo de Trabalho, *op. cit.*, p. 10).

[45] NARAYANAN, Arvind; FELTEN, Edward; *op. cit.*, p. 3.

caso mediático da identificação de utilizadores da plataforma Netflix[46], o qual rapidamente se tornou o exemplo paradigmático da fragilidade das técnicas de anonimização perante a abundância de dados publicamente disponíveis[47].

Como mecanismo de prevenção, e para efeitos de colmatar o risco de identificação através da associação com outras bases de dados, sugere-se a implementação de medidas integradas na família de técnicas de anonimização enquadradas na generalização. Ao aplicar medidas particulares de agregação, como o K-anonimato[48], estas abordagens teriam o condão de generalizar ou diluir os atributos dos titulares dos dados, complexificando a identificação, mesmo com o recurso a outras fontes de informação. No entanto, existem investigações que reportam que as garantias que outrora pudessem ter sido prestadas pela generalização (no âmbito de uma conjetura na qual prevaleciam bases de dados de cariz unidimensional) deixou de se verificar no panorama atual do tratamento de dados pessoais. Uma vez que proeminam hoje as bases de dados de cariz complexo[49], refere-se que a abundância de informação publicamente

---

[46] NARAYANAN, Arvind; SHMATIKOV, Vitaly. "Robust de-anonymization of large sparse datasets (How to break anonymity of the Netflix prize dataset)", *IEEE Symposium on Security and Privacy, 2008, p. 111-125,* Disponível em: https://www.semanticscholar.org/paper/How-To-Break-Anonymity-of-the-Netflix-Prize-Dataset-Narayanan-Shmatikov/56116e8ce3f57bec578ac60f6d68333aea5af59e, acedido a: 21.04.2019.

[47] Em 2 de outubro de 2006, a provedora global de filmes e séries de televisão via *streaming* Netflix promoveu um concurso tendente à elaboração de um algoritmo que promovesse a sua recomendação de filmes. Para efeitos operativos, a empresa publicou informação referente a 100.480.507 avaliações realizadas por 480.189 dos seus subscritores no período compreendido entre dezembro de 1999 e dezembro de 2005. O processo de anonimização foi feito através da remoção de informação passível de identificação direta dos subscritores. Contudo, utilizando como fonte de informação adicional os dados publicamente disponíveis no site *Internet Movie Database* (conhecido como "*IMDb*"), demonstrou-se que era possível estabelecer uma conexão entre os registos e os subscritores, contornando os esforços de anonimização.

[48] As técnicas de agregação, nas quais o k-anonimato se insere, procuram impedir que o titular seja identificado através da sua agregação com outros sujeitos que partilham um elemento geral comum. A particularidade do k-anonimato reside no facto de que pretende assegurar que nesse agrupamento exista, pelo menos, um determinado número de titulares, no sentido de evitar a individualização de um sujeito em particular.

[49] ZIBUSCHKA, Jan, *et al., op. cit.*, p. 4.

disponível coloca em causa a proteção garantida pela anonimização[50], mesmo no caso de bases de dados anonimizadas com uma elevada taxa de generalização[51].

A conclusão que devemos retirar desta conjetura é que a abundância de informações livremente disponibilizadas (incluindo as que são promovidas pelos próprios titulares no âmbito do ecossistema digital) promove, necessariamente, a individualização da população com o aumento dos atributos disponíveis[52]. Como tal, a probabilidade de identificar um indivíduo é alta, o que acaba por comprometer o nível de proteção assegurado pela redução da individualidade dos dados. Perante estas limitações, a doutrina vem defender que a anonimização eficiente de determinado tipo de dados pessoais, mais do que uma dificuldade técnica, representa-se como uma efetiva impossibilidade prática[53], chegando-se a falar de uma verdadeira quebra de confiança na anonimização enquanto meio de tutela de dados pessoais.

2.2.2. Influência de elementos de carácter intrínseco: O impacto da natureza dos dados na viabilidade da anonimização

Embora concebamos que uma das principais limitações do processo de anonimização poderá ser encontrada no contexto no qual este se

---

[50] ROCHER, Luc, *et al., op. cit.*, p. 10.

[51] Ainda assim, porém, encontramos argumentos que defendem que, desde que seja realizada uma correta implementação de técnicas de anonimização, mesmo quando se logra reverter o respetivo processo, permanece sempre um determinado grau de incerteza. Assente neste nível de dúvida, alega-se que seria possível ao titular da informação sonegar a sua representação nos dados evidenciados, o que, em última análise, poderia representar um reforço na tutela do sujeito identificado. Contudo, estudos recentes vêm rebater este tipo de argumentação, tendo-se logrado desenvolver modelos que apontam para a existência de um alto grau de probabilidade de que o resultado de tarefas de reversão sejam representações corretas dos dados reais (ROCHER, Luc, *et al., op. cit.*, p. 1-5).

[52] GOLLE, Philippe. *Revisiting the Uniqueness of Simple Demographics in the US Population*. Palo Alto Research Center, 2006, p. 1, disponível em: https://www.privacylives.com/wp-content/uploads/2010/01/golle-reidentification-deanonymization-2006.pdf., acedido a: 14.05.2018.

[53] ZIBUSCHKA, Jan. *et al*, *op. cit.*, p.71.

enquadra[54], a verdade é que a avaliação casuística exigida pelo critério da razoabilidade demanda, claramente, uma análise cuidada do tipo de dados que se pretende anonimizar. Uma vez que cada técnica de anonimização possui debilidades que lhe são inerentes, e tendo em conta que não existe um caminho pré-determinado para descortinar objetivamente quais as medidas exatas a implementar, a sua eventual adequação ao caso concreto dependerá sempre das particularidades dos dados em causa[55].

Aliás, do nosso ponto de vista, julgamos poder encontrar uma relação direta entre a evolução destas especificidades ao longo do tempo e o declínio do nível de eficiência das técnicas de anonimização. Nomeadamente, num contexto de limitada capacidade de processamento de informação, como o que se verificava outrora, no qual predominava um tratamento desprovido da complexidade multidimensional da atualidade, as técnicas tradicionais de anonimização poderiam, efetivamente, representar maiores garantias, uma vez que seria mais simples proceder à generalização dos seus atributos. Contudo, a atualidade é caracterizada pela proliferação das bases de dados de cariz complexo, maioritariamente compostas pela denominada *high-dimentional data*[56], e no âmbito da qual se evidencia um claro limite quanto à medida de generalização de que determinados dados podem ser alvo sem que se comprometa gravemente a sua integridade.

Devido à elevada taxa de individualização[57] das informações tratadas hoje em dia, investigações recentes procuram demostrar que a

---

[54] ELLIOT, Mark. *et al*, *op. cit.*, p. 67.

[55] Grupo de Trabalho, *op. cit.*, p. 28.

[56] Na sua essência, este conceito refere-se às informações constituídas por uma grande quantidade de elementos identificadores para cada indivíduo, de tal forma que os registos individuais possuem uma grande probabilidade de serem únicos e diferentes dos demais registos (neste sentido, ZIBUSCHKA, Jan, *et al.*, *op. cit.*, p. 4).

[57] A dificuldade na anonimização deste tipo de dados está intimamente ligada aos traços característicos do comportamento humano. Nomeadamente, num conjunto de dados anónimos, os movimentos pendulares diários permitem assumir, com um certo nível de segurança, que determinados pontos representam o local trabalho e a residência de uma dada pessoa. Como tal, os dados de localização têm uma grande singularidade. Em determinados estudos, chega-se a referir que perto de 95% de rastreios, compostos por quatro locais, tem um carácter único (neste sentido, MONTJOYE, Yves-Alexandre; HIDALGO, César A.; VERLEYSEN, Michel; BLONDEL, Vincent. "Unique in the Crowd: The privacy bounds of human mobility". *Scientific Reports*, v. 3, n.º 1376, p. 1-5, disponível em: https://www.nature.com/articles/srep01376, acedido a: 21.05.2019).

anonimização de determinados dados de cariz complexo (como seja a generalidade dos dados referentes a rastreamentos de dispositivos de comunicação[58], os dados referentes às redes sociais[59], determinados registos médicos ou dados comportamentais em geral[60]) poderá consubstanciar um desafio efetivamente intransponível[61]. Por outro lado, se considerarmos ainda a mencionada possibilidade de conjugação deste tipo de dados com o crescente acervo de informação livremente disponível, facilmente se antevê a provável extrapolação do risco de identificação.[62]

Ou seja, além de representar um elemento de complexificação da tarefa exigida ao responsável pelo tratamento, a consideração das particularidades dos dados a tratar reveste-se de particular importância, uma vez que, em última análise, poderá comprometer a própria viabilidade efetiva da anonimização.

2.3. O carácter antagónico das finalidades visadas pelo processo de anonimização

2.3.1. A minimização do risco residual

À parte de quaisquer considerações relativas ao sucesso efetivo deste processo, a anonimização traduz-se, acima de tudo, num mecanismo de tutela do direito fundamental à proteção de dados pessoais. Nesta medida, a sua finalidade principal concretiza-se num esforço de pendor preventivo que se concentra na proteção efetiva dos direitos do titular.

Conforme esquematizado pelo Grupo de Trabalho, esta finalidade seria lograda se a anonimização conseguisse impedir que, com base em certos

---

[58] MONTJOYE, Yves-Alexandre, *et al.*, *op. cit.*

[59] UGANDER, Johan; KARRER, Brian; BACKSTROM, Lars; MARLOW, Cameron. "The Anatomy of the Facebook Social Graph.", 2011, disponível em: https://arxiv.org/abs/1111.4503, acedido a: 22.10.2019.

[60] ZIBUSCHKA, Jan, et al., op. cit., p. 76.

[61] NARAYANAN, Arvind. *et al*, *op. cit.*, p. 1.

[62] NARAYANAN, Arvind. *et al*, *op. cit.*, p.2.

dados, fosse possível qualquer identificação[63], ligação[64] ou inferência[65] em relação aos respetivos titulares. No entanto, e conforme resulta patente dos considerandos do RGPD, não se exige que essa finalidade seja lograda em termos absolutos. Ou seja, não é exigível ao responsável que o risco de identificação, ligação ou inferência seja reduzido à nulidade.

Partindo destes pressupostos, o legislador europeu parece compreender e aceitar o facto de que não existem técnicas de anonimização que disponibilizem uma resposta cabalmente satisfatória a esta questão[66]. Embora lhes seja reconhecido um grau variável de adequação consoante as circunstâncias concretas, as investigações levadas a cabo nesta área revelam sistematicamente que nenhuma técnica é, por si só, desprovida de lacunas[67]. É partindo desta perspetiva que a legislação europeia abraça um conceito lato de anonimização, o qual se encontra apto a compreender e tolerar a permanência de um fator de risco inerente ao dado anonimizado[68]. Trata-se de um risco de carácter necessariamente residual, cuja eliminação representaria uma condição inexigível ao responsável pelo tratamento, uma vez que o seu aproveitamento para efeitos de identificação seria desrazoável (tendo em conta os meios disponíveis).

---

[63] Nos termos do referido pelo Grupo de Trabalho, a identificação ocorrerá quando seja possível "isolar alguns ou todos os registos que identifiquem uma pessoa num conjunto de dados" (Grupo de Trabalho, *op. cit.* p. 3). Em concreto, existirá o risco de identificação quando for possível individualizar o titular dos dados, mesmo após a conclusão do processo de anonimização.

[64] O risco de ligação corresponde à possibilidade de se determinar uma relação entre os registos referentes a um mesmo titular. Relativamente a este aspeto, o Grupo de Trabalho vem alertar para o facto de que uma técnica poderá ser efetiva contra a identificação, mas, ainda assim, ser ineficaz contra o risco de ligação. Isto sucederá quando uma determinada técnica de anonimização logre coartar o destacamento do indivíduo num determinado grupo, porém, seja ainda possível determinar que um grupo de dados se refere a um mesmo grupo de pessoas (Grupo de Trabalho, *op. cit.* p. 3).

[65] Ainda nos termos do avançado pelo Grupo de Trabalho, fundamentalmente, a inferência refere-se à "possibilidade de deduzir (...) o valor de um atributo a partir dos valores de um conjunto de outros atributos". Devido à natureza deste conceito, a possibilidade da sua verificação estará sempre pendente de uma avaliação de probabilidade na eventual concretização efetiva das informações inferidas (Grupo de Trabalho, *op. cit.* p. 3).

[66] Grupo de Trabalho, op. cit., p. 9.

[67] *Ibid.*, p. 13.

[68] *Ibid.*, p. 7.

A contraposição necessária desta construção implica que o risco não possa exceder o carácter residual. Caso tal se verifique, o processo de anonimização deverá ser considerado débil, uma vez que o respetivo risco ultrapassa o critério da razoabilidade, pelo que os dados produzidos pelo mesmo não possuirão a natureza de dado anonimizado.

Desonerado da responsabilidade da eliminação absoluta do risco, ao responsável pelo tratamento é, porém, exigido que encete todos os meios razoavelmente disponíveis para a sua minimização. É sobre esta premissa que o responsável pelo tratamento deverá avaliar a conjugação das técnicas de anonimização a utilizar[69]. Cada vez mais, a doutrina tem evidenciado que uma das falhas correntes das práticas atuais de anonimização se encontra na aplicação destas técnicas sem ter em conta a sua efetividade no caso concreto (ou mesmo o impacto que o procedimento possa ter nos próprios dados)[70]. Uma vez que as diversas técnicas de anonimização possuem benefícios e limitações próprias[71], torna-se essencial realizar uma ponderação casuística que se centre na efetividade concreta das técnicas de anonimização, avaliando a pertinência de eventuais combinações de técnicas no sentido da sua complementação[72].

Devido à natureza dinâmica dos elementos contextuais que influem sobre a efetividade da anonimização, devemos salientar que a minimização do risco residual não é uma condição estática e exige uma ponderação contínua de carácter preventivo. O responsável pelo tratamento deverá assegurar-se que o nível de risco se mantém, efetivamente, residual ao longo do tempo[73].

2.3.2. A maximização da utilidade residual

Sem nos desviarmos das premissas alcançadas até este ponto, e muito embora a proteção dos direitos do titular dos dados assuma um papel

---

[69] PANG, Ruoming; ALLMAN, Mark; PAXSON, Vern; LEE, Jason. "The Devil and Packet Trace Anonymization". *ACM SIGCOMM Computer Communication Review*, 36. p. 29-38, disponível em: https://www.icir.org/enterprise-tracing/devil-ccr-jan06.pdf, acedido a: 22.08.2019.

[70] NALDI, Maurizio, *op. cit.*

[71] Grupo de Trabalho, op. cit., p. 26.

[72] *Ibid.*

[73] NARAYANAN, Arvind. et al, op. cit., p. 5.

epicêntrico nesta matéria, devemos manter presente que a minimização do risco residual não se afigura como o único objetivo do processo de anonimização[74].

Adensando a complexidade que reveste este processo, através de uma abordagem eminentemente pragmática, a doutrina tem procurado realçar que a minimização do risco representa apenas um dos vértices deste problema[75]. Como é obvio, todo o tratamento de dados pessoais que vise a sua anonimização, além da proteção dos titulares, tem sempre em vista a sua futura utilização[76]. No sentido de promover essa finalidade, além da existência do risco residual, o responsável pelo tratamento deverá considerar também o nível de utilidade que os dados conservarão após o processo de anonimização[77]. É precisamente nesta dicotomia que se encontra o paradoxo que complexifica o processo de anonimização no plano das suas finalidades[78].

Conceptualmente, e para efeitos de exposição, seria possível conceber um processo de anonimização *stricto sensu* que, ao deturpar completamente a informação, lograsse impedir categoricamente a recondução de quaisquer valores aos respetivos titulares e vice-versa. Evidentemente, este ponto representaria um nível máximo de anonimização e, consequentemente, de proteção do titular dos dados. Tal adulteração, porém, traduzir-se-ia num nível mínimo (ou nulo) de utilidade dos dados tratados, uma vez que esse tratamento colocaria em causa a fidelidade – e consequente utilidade – dos mesmos[79]. Inversamente, dados originais, ou sujeitos a um procedimento débil de anonimização, embora representem um alto nível de utilidade (uma vez que não se afastariam dos valores originais) traduziriam um nível de anonimização mínimo ou inexistente[80]. Perante este contraste, e se tivermos presente que não é incomum que o principal

---

[74] BAMBAUER, Jane; MURALIDHAR, Krishnamurty; SARATHY, Rathindra. "Fool's Gold: An Illustrated Critique of Differential Privacy". *Vanderbilt Journal of Entertainment and Technology Law*, 16, n.º 4, 2014, p. 701-755, disponível em: http://www.jetlaw.org/wp-content/uploads/2014/06/Bambauer_Final.pdf, acedido a: 26.08.2019.

[75] NALDI, Maurizio, *op. cit.*

[76] ELLIOT, Mark. *et al, op. cit.*, p. 18.

[77] Grupo de Trabalho, *op. cit.*, p. 4.

[78] ZIBUSCHKA, Jan, *et al.*, *op. cit.*, p. 71.

[79] *Ibid.*, p. 55.

[80] ZIBUSCHKA, Jan. *et al*, *op. cit.*, p.76.

elemento de risco de uma determinada base de dados seja precisamente o fator de maior utilidade, facilmente se antevê a dificuldade que esta conformação representa[81].

Torna-se evidente, assim, que esta correlação demanda um equilíbrio próprio, cuja ponderação deverá assumir um papel fulcral no processo de anonimização[82], pois, da sua frustração, poderá decorrer a futilidade de todo o processo, quer seja pela inutilização dos dados tratados ou pela ineficácia das medidas implementadas.

A natureza pragmática da questão condiciona a sua avaliação a um critério necessariamente casuístico e virado para as finalidades específicas do caso, porém, para efeitos de esquematização da lógica que deverá orientar esta consideração, podemos afirmar que, independentemente das particularidades da questão em concreto, o responsável deverá sempre procurar almejar um ponto de equilíbrio[83]. Uma vez que o incremento num dos ideais comporta, necessariamente, um prejuízo para o outro, entendemos que a conformação de uma estratégia de anonimização torna essencial procurar um ponto que, por um lado, procure evitar ao máximo os riscos identificados *supra* e, por outro, procure ainda assegurar um máximo de utilidade. Apesar de que, na prática, este patamar seja virtualmente impossível de antever, em termos teóricos, será possível conceber, em torno deste ponto, um espaço de razoabilidade a nível do risco e da utilidade que logre satisfazer o conceito *lato sensu* consagrado no âmbito da legislação europeia.

Não deixa de ser evidente, porém, que esta avaliação estará sempre revestida de uma índole necessariamente subjetiva e casuística que lhe confere uma dimensão de fragilidade que, por sinal, desafia a efetividade da maior parte das técnicas de anonimização. Ilustrativamente, poderíamos indicar a implementação de mecanismos baseados na

---

[81] ELLIOT, Mark. *et al, op. cit.*, p. 41.

[82] NALDI, Maurizio, op. cit.

[83] Para efeitos de ilustração do raciocínio vertido, fazemos referência à lei da eficiência de Pareto, desenvolvida pelo economista Vilfredo Frederico Damaso Pareto, na sua obra Cours d`Économie Politique de 1897. Em termos simplistas, este conceito pretende representar uma conjetura na qual se procede à alocação de recursos de forma a que não se vislumbre uma outra organização de recursos viável que possa promover nenhum dos interesses sem que, em contrapartida, se implique um prejuízo para o outro.

privacidade diferencial[84] como a procura de soluções de compromisso para esta relação dicotómica[85]. Tratando-se de um método dinâmico de particularização do nível de distorção dos dados ao caso específico da sua consulta, este mecanismo providenciaria uma relação equilibrada entre o risco e a utilidade. Contudo, a verdade é que não presta uma resposta definitiva a esta questão. Importa recordar que, além das limitações que lhe são comumente imputadas[86], no seu esforço por prestar uma solução de equilíbrio entre as duas finalidades, esta técnica acaba por sacrificar a natureza anónima dos dados que produz. Uma vez que favorece a preservação de uma versão original dos dados, mesmo após a implementação deste mecanismo, mantém-se uma via concreta para a identificação dos titulares. Como tal, os dados sujeitos a esta técnica não poderão ser considerados aquém do âmbito material do RGPD e, nesta medida, pelo menos na iteração atual dos contornos desta técnica, ainda não logra representar uma resposta definitiva à questão da proteção de dados mediante a anonimização.

Indo além das limitações próprias das diversas técnicas de anonimização, devemos também realçar que a natureza da informação poderá dificultar a procura por uma relação de equilíbrio. Torna-se hoje comum

---

[84] WOOD, Alexandra; ALTMAN, Micah, BEMBENE, Aaron; BUN, Mark; GABOARDI, Marco; HONAKER, James; NISSIM, Kobbi; O´BRIEN, David; STEINKE, Thomas; VADHAN, Salil. "Differential privacy: A primer for a non-technical audience.", *Vanderbilt Journal of Entertainment & Technology Law*, 21, n.º 1, 2018, p. 209-275, disponível em: http://privacytools.seas.harvard.edu/files/privacytools/files/pedagogical-document-dp_0.pdf., acedido a: 26.08.2019.

[85] Considerada como um passo em frente em relação à proteção de dados e à tutela da privacidade, esta técnica desenvolve as soluções baseadas na adição de ruído, procurando, ao mesmo tempo, promover a preservação da utilidade dos dados. Em lugar de imbuir uma medida pré-determinada de distorção à informação, esta técnica conserva uma cópia dos dados originais e se limita a aplicar um certo nível de ruído apenas aos resultados visualizados pelo sujeito que procede à consulta da base de dados. Em suma, trata-se de um mecanismo particularizado e especializado ao terceiro que a consulta, permitindo uma contínua adaptação à necessidade de adição de ruído.

[86] Tal como sugere o grupo de trabalho, através do acesso contínuo à base de dados, eventualmente por um conjunto diferenciado de indivíduos, seria conceptualmente possível retirar informação que permita a identificação, ligação ou inferência sobre alguns dos titulares. Muito embora o enquadramento dos dados como dados pessoais lograsse manter as garantias aos titulares, não se poderá considerar que estejam ultrapassados os riscos elencados *supra*.

encontrar contributos científicos que evidenciam que, no âmbito do tratamento de determinado tipo de dados, a minimização do risco a um nível razoavelmente satisfatório implica sempre uma desmedida distorção dos mesmos, prejudicando incontornavelmente a sua utilidade[87]. Principalmente no âmbito de dados complexos, nos quais existe um número considerável de quase-identificadores, a limitação do nível de generalização que poderá ser efetuada sem comprometer a utilidade dos dados é alta[88]. Nesta medida, chega-se mesmo a referir que, em relação a determinados tipos de dados, um processo de anonimização com um alto índice de utilidade e segurança é uma impossibilidade fáctica[89].

É nas complexas nuances deste equilíbrio que se desenvolvem as principais dúvidas quanto à virtude da anonimização enquanto meio de proteção de dados pessoais[90]. Na medida em que assenta na premissa questionável da possibilidade de obter dados úteis e, ao mesmo tempo, verdadeiramente anonimizados[91], pondera-se a necessidade efetiva de uma mudança de paradigma que se afaste da anonimização e se foque na transparência e numa lógica de responsabilidade[92].

Considerações finais

Tal como procurámos antecipar na introdução do presente texto, muito embora a proteção de dados pessoais se revele como um direito fundamental amplamente reconhecido em termos globais (embora em diferentes níveis), os seus contornos são constantemente moldados e desafiados pelo desenvolvimento da tecnologia da informação[93], a qual medra num ritmo distinto do que a capacidade de adaptação das soluções jurídicas, representando um especial desafio para os esforços regulatórios.

---

87 BAMBAUER, Jane. *et al, op. cit.*, p 703.
88 NARAYANAN, Arvind. *et al, op. cit.*, p. 7.
89 MURAKAMI, Takao. *et al, op. cit.*
90 ELLIOT, Mark. *et al, op. cit.*, p. 18.
91 OHM, Paul., *op. cit.*, p. 1704.
92 ZIBUSCHKA, Jan, et al., op. cit., p 79.
93 *Ibid.*, p.71.

Assente nas considerações tecidas até este momento, entendemos que a viabilidade prática de um processo de anonimização está pendente de um equilíbrio, necessariamente ténue, entre a utilidade e o risco residual[94] que, em última análise, se encontra nas mãos do próprio responsável pelo tratamento e cuja resiliência depende da ponderação de uma universalidade de fatores que lhe são alheios[95], como a conjetura contemporânea dos recursos tecnológicos e a disponibilidade de informação que os próprios titulares continuamente disseminam. Permitimo-nos, assim, questionar a viabilidade de um regime assente numa ponderação de razoabilidade de meios. Já de si um conceito indeterminado, este critério poderá consagrar-se como uma realidade indeterminável se considerarmos o seu efetivo enquadramento prático.

Por um lado, o processamento de informação encontra-se enquadrado num contexto dinâmico que promove uma constante evolução de meios e recursos que impossibilita a antevisão do que poderá ser considerado concebível a curtíssimo prazo. Por outro lado, conforme também observámos, a doutrina revela como uma crescente fonte de informações publicamente disponibilizadas compromete a efetividade das técnicas de anonimização. A questão é que a dificuldade deste processo não se traduz necessariamente numa maior proteção do titular. O que os casos de estudo têm demostrado é que a complexidade da sua ponderação tem resultado na produção de dados ilusoriamente anonimizados que promove a vulnerabilidade dos titulares.

Assim sendo, entendemos que, por se compaginar na realidade volátil e mutável do ecossistema digital e tecnológico, o critério da razoabilidade de meios poderá encontrar-se vazio de conteúdo, uma vez que poderá ser impossível determinar o que é, num dado momento, efetivamente razoável e, em última análise, poderá contribuir para o desenvolvimento de um ambiente de incerteza por parte dos responsáveis pelo tratamento e de desconfiança por parte dos titulares dos dados.

Transpondo a nossa objetiva ao regime consagrado pela União aos dados anonimizados, observamos como estas considerações nos levam a suspeitar que na delicada ponderação de razoabilidade se encontra pendente a possibilidade de sonegação dos mais básicos direitos e deveres

---

[94] ELLIOT, Mark. *et al*, *op. cit.*, p. 19.

[95] *Ibid.*, p. 16.

decorrentes do RGPD. Neste sentido, julgamos que seria, efetivamente, defensável uma mudança de paradigma a nível da tutela dos dados anónimos ou do próprio conceito de anonimização[96].

Conforme é nosso entendimento, o enquadramento legislativo europeu referente à proteção de dados encara o resultado do processo de anonimização de dados pessoais em termos equiparados ao efetivo apagamento dos mesmos. Salvo exceções técnicas e recursos de recuperação de informação, em princípio, um dado apagado estará objetivamente inacessível. No caso da anonimização, como vimos, a reidentificação do titular não é uma impossibilidade objetiva, na medida em que está pendente de um juízo de razoabilidade eminentemente subjetivo. Esta fragilidade, bem como a dimensão omnipresente de risco que inerentemente comporta, não é coerente com esta equiparação. Como tal, entendemos que a consagração de um regime jurídico como o que se verifica à luz da atual legislação europeia apenas se justificaria no contexto de uma conceção absoluta ou *strito sensu* de anonimização[97].

Ao aceitar-se a impossibilidade fáctica de tal realidade, consideraríamos coerente revestir o processo de anonimização *lato sensu* de um núcleo básico de garantias. A título meramente exemplificativo, e admitindo a necessidade de maturação da presente construção, se, efetivamente, o processo de anonimização comporta sempre um risco para o respetivo titular, não se poderia aceitar a conclusão de que este tratamento pudesse ter lugar sem o assentimento ou intervenção deste último[98].

---

[96] ZIBUSCHKA, Jan, *et al., op. cit.,* p. 72.

[97] *Ibid*., p. 78.

[98] Com efeito, e uma vez que os dados pessoais permanecerão sob a tutela do RGPD até que sejam sujeitos ao processo de anonimização, este tratamento (conforme definido pelo n.º 2 do art. 4.º do RGPD) deverá ser realizado em observância do princípio da licitude [consagrado na alínea a) do n.º 1 do art. 5.º do RGPD]. Contudo, o responsável poderá invocar qualquer um dos fundamentos de licitude previstos nas várias alíneas do n.º 1 do art. 6 do RGPD, não estando limitado ao consentimento. Mais, se observarmos ainda que este tratamento não contraria necessariamente os vetores do princípio da limitação das finalidades [consagrado na alínea b) do n.º 1 do art. 5.º do RGPD] – uma vez que o mesmo, em princípio, poderá ser considerado compatível com a finalidade para a qual os respetivos dados foram originalmente recolhidos –, torna-se evidente que o processo de anonimização poderá tecnicamente desenrolar-se de forma alheia aos titulares dos dados (neste sentido Grupo de Trabalho, op. cit., p. 8).

Muito embora, em termos conceptuais, o processo de anonimização vise representar uma mais-valia para a proteção do titular dos dados, a existência de um nível de risco constante (bem como o facto de que este processo poderá coartar os direitos e deveres constantes do RGPD), implica que o interesse legítimo do responsável pelo tratamento não possa ser considerado, *à priori*, elemento suficiente para encetar a anonimização[99]. Em nome da transparência, e para efeitos da proteção dos direitos do titular dos dados, consideramos que o processo de anonimização careceria de comprovação de que o mesmo tem consciência do nível de risco e o aceita como tal. Ou seja, deverá reconhecer-se a fragilidade da anonimização e promover a implementação de uma correta política de transparência e consentimento[100].

Nesta mesma linha de raciocínio, seria também defensável a concretização do princípio da necessidade ao processo de anonimização. Na prática, atendendo à omnipresença do referido fator de risco, o responsável pelo tratamento não deverá apenas provar que considera as expectativas dos titulares neste processo, mas também demostrar a concreta necessidade de optar pela anonimização e conservação dos dados, em lugar de proceder à efetiva eliminação dos mesmos.

Seria, nesta medida, de ponderar, como tem acontecido na literatura recente[101], a possibilidade de repensar a abordagem consagrada à

---

[99] Por sinal, deverá salientar-se que, muito embora o consentimento pudesse ser equacionado como fundamento por excelência em nome da transparência, a doutrina especializada tem defendido que, para efeitos de anonimização, o consentimento poderá revelar-se contraproducente (EMAM, Khaled El; HINTZE, Mike. *Does anonymization or de-identification require consent under the GDPR?,* 2019, disponível em: https://iapp.org/news/a/does-anonymization--or-de-identification-require-consent-under-the-gdpr, acedido a: 14.02.2019). Nos termos deste entendimento, além de ser difícil de requerer no âmbito do processamento de *big data* ou no contexto da aprendizagem automática (*machine learning*), o consentimento do titular, para tais efeitos, poderia corromper determinados resultados devido à sua índole tendenciosa. Conforme defende esta abordagem, os titulares que aderem ao tratamento, possuem um perfil tendencialmente semelhante, o que, ao excluir os que não consentiriam, poderia subverter a fidelidade dos dados. Como tal, defende-se que, no contexto da anonimização de dados, a legitimação da anonimização através do interesse legítimo [nos termos da alínea f) do n.º 1 do art. 6.º do RGPD] poderia revelar-se mais pertinente do que a concentração no consentimento do titular.

[100] ZIBUSCHKA, Jan, *et al., op. cit.,* p. 79.

[101] *Ibid.*, p. 72.

anonimização, revestindo-a de uma tutela coerente com os objetivos consagrados no âmbito da estratégia europeia para o mercado único europeu que, no seio dos seus principais baluartes, consagra o desenvolvimento da confiança no quadro do ecossistema digital.