As *Smart Cities* e a Privacidade: o critério legal para a anonimização de dados agregados

JOANA DINIZ DE FIGUEIREDO[*]

Resumo: O presente estudo visa estudar a técnica de agregação de um ponto de vista legal de forma a determinarmos se os dados obtidos são anonimizados ou pessoais, e caso sejam pessoais, a partir de que momento é que deixam de ser. Esta conclusão é essencial para compreender se o tratamento de dados agregados no âmbito das Smart Cities encontra-se abrangido pelo RGPD1 . Para alcançar esta solução, analisámos o contexto das Smart Cities, o conceito de dados agregados e de anonimização, o modelo de agregação e concluímos através da definição de um critério que permite aferir o momento a partir do qual os dados agregados são dados anónimos à luz da lei aplicável.

***Palavras-chave**: cidades inteligentes; anonimização; dados agregados; privacidade; princípio do "privacy by design".*

Abstract: The aim of this study is to analyze the aggregation technique from a legal perspective to determine whether the obtained data is anonymized data or personal data. Moreover, it intends to ascertain, in this second case, when will it cease to be personal data. Such conclusion is essential to understand whether the processing of aggregated data within the scope of Smart Cities is covered by the GDPR. To achieve this solution, we analyzed the context of Smart Cities, the concept of aggregated data, and the concepts of technical

[*] Advogada e Consultora na área de Proteção de Dados, Direito Digital e TMT. Licenciada em Direito e Mestre em Direito Forense pela Universidade Católica Portuguesa, Faculdade de Lisboa.

and legal anonymization. Furthermore, we studied the aggregation model within the scope of Smart Cities and concluded by defining a criterion that allows us to determine when the aggregated data is anonymous according to the applicable law.

Keywords*: smart cities, anonymization, aggregated data, privacy, privacy by design.*

1. Introdução

O conceito de "Smart Cities" ou cidades inteligentes tem vindo a ganhar uma enorme importância nos últimos anos como um meio para fazer face aos desafios que a nossa sociedade enfrenta. A necessidade de aumentar a qualidade de vida dos cidadãos, a eficiência e a qualidade dos serviços, a adequação das medidas e serviços aos cidadãos e a sustentabilidade são alguns dos objetivos pretendidos nos dias de hoje e que despoletam os mais variados estudos e investigações.

Os projetos de Smart Cities têm como pedra de toque a utilização de novas tecnologias de informação que implicam a recolha de dados com grande precisão, qualidade e em grande quantidade. Tecnologias como a utilização de drones, impressoras 3D, Blockchain, Big Data, Internet of Things ("IoT"), Gamification, Smart Cars, Cloud e Inteligência Artificial ("IA") são consideradas o combustível das cidades inteligentes. Contudo, conforme analisaremos no presente estudo, esta recolha massiva de dados pessoais dos cidadãos poderá colocar variados desafios à privacidade e à proteção dos dados dos mesmos. Assim sendo, é necessária a procura de soluções que permitam que os municípios maximizem os benefícios desta recolha de dados com o total respeito pelos direitos, liberdades e garantias dos titulares dos dados.

Nesta senda, o presente estudo visa analisar uma solução que permita alcançar um equilíbrio entre a maximização da utilidade dos dados e o respeito pela proteção de dados dos cidadãos, em particular, através da técnica da agregação de dados. Ainda que muitas vezes seja considerado um método de anonimização de dados, vamos analisar se os dados resultantes da aplicação

desta técnica são dados pessoais ou se são dados anónimos para efeitos da legislação aplicável em matéria de proteção de dados.

A necessidade do presente estudo prendeu-se com a constatação da essencialidade do respeito pela privacidade dos cidadãos e da transparência relativamente aos projetos de Smart Cities. Apenas através de um projeto totalmente transparente podemos alcançar uma verdadeira proximidade com os cidadãos e a confiança dos mesmos. Contudo, estas considerações devem ser tomadas ad initium, alinhadas com os objetivos[1] do Regulamento Geral de Proteção de Dados (doravante designado por "RGPD" e "Regulamento")[2] e com o Princípio do Privacy by Design[3].

2. Smart Cities

2.1. Conceito

Como resultado do crescimento da população nas metrópoles, da redução dos recursos disponíveis e do desenvolvimento exponencial da tecnologia, as cidades têm vindo a ser estimuladas a encontrar novos métodos e soluções para alcançar uma maior eficiência, sustentabilidade, resiliência e incrementar a qualidade de vida dos cidadãos[4]. Através das cidades inteligentes (Smart Cities) procura-se responder a estes principais problemas enfrentados pelos espaços urbanos[5].

---

1 "(…) na proposta de Regulamento, a Comissão enuncia três grandes objetivos: 1- Permitir o desenvolvimento da economia digital; 2- Permitir que as pessoas singulares controlem os seus próprios dados; 3- Reforçar a segurança jurídica e prática para os operadores económicos e as entidades públicas." in Berberan Santos, Sofia e Gabriel, João, *Regulamento Geral Sobre a Proteção de Dados, Legislação e Algumas notas*, 3.ª Edição, Edição GPA Academy (2020), 20.

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.

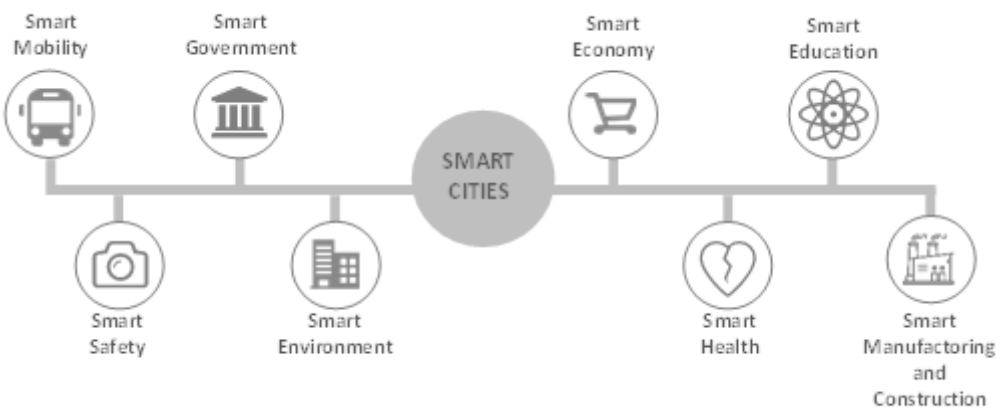
3 Cfr. art. 25.º do Regulamento Geral de Proteção de Dados.

4 "*Smart cities have emerged in this context as an answer to the growing problems of unsustainable urban expansion, growing inequality, climate change, and insecurity. Smart cities are urban centres that harness technologies such as big data, algorithms, and Internet of Things ('IoT') to enhance innovation and urban competitiveness.*" in Ranchordás, Sofia e Klop, Abram "Data-driven Regulation and Governance in Smart Cities", *Research Handbook in Data Science and Law*, 2 (2018).

5 "The vision of *''Smart Cities"* is the urban center of the future, made safe, secure

Assim sendo, as cidades consideram-se inteligentes quando o investimento em capital humano, infraestruturas tradicionais e tecnologias disruptivas estimulam o crescimento económico sustentável e a melhoria significativa da qualidade de vida, com uma gestão inteligente dos recursos naturais[6]. As cidades inteligentes são contruídas através de um conjunto de soluções ditas "inteligentes" nos diversos setores de atividade da nossa sociedade, onde são aplicadas tecnologias no sentido de incrementar a eficiência dos serviços e das tomadas de decisão (*vide* figura 1).

Figura 1 – Exemplos de Setores Abrangidos pelas Soluções Inteligentes



Fonte: Elaboração própria

A utilização das novas tecnologias para alcançar cidades sustentáveis, conectadas e otimizadas já não representa um futuro longínquo, sendo concebido como o meio principal para alcançar os objetivos referidos. Tecnologias como a utilização de drones, impressoras 3D, *Blockchain, Big Data, Internet of*

environmentally green, and efficient because all structures--whether for power, water, transportation, etc. are designed, constructed, and maintained making use of advanced, integrated materials, sensors, electronics, and networks which are interfaced with computerized systems comprised of databases, tracking, and decision-making algorithms." in Robert E. Hall/B. Boweman/J. Braverman/J. Taylor/H.Todosow/U.von Wimmersperg, The Vision of a Smart City, 2nd International Life Extension Technology Workshop, 1 (2000).

6 Deloitte, Smart Cities, How rapid advances in technology are reshaping our economy and society, version 1.0, (Nov.- 2015). Acessível em: <https://www2.deloitte.com/tr/en/pages/public-sector/articles/smart-cities.html> (consultado a 22 de novembro de 2020).

Things ("IoT"), Gamification, Smart Cars, Cloud e Inteligência Artificial ("IA") são consideradas o combustível das cidades inteligentes. Como foi apresentado por um relatório da Deloitte sobre *Smart Cities* "*[s]mart cities exist on the intersection of digital technology, disruptive innovation and urban environments. They are an exciting place to work and live and the breeding ground for new ideas*"[7]. As Tecnologias da Informação e Comunicação ("TIC") são desenvolvidas a uma velocidade sem precedentes ao longo dos últimos anos e a conjunção destas com os ambientes urbanos está a criar ambientes urbanos bastante diferentes do experimentado até agora.

Neste contexto, os projetos e programas de Smart Cities começam a emergir por todo o mundo como um novo paradigma e como a resposta adequada aos desafios da nossa sociedade.

A *PricewaterhouseCoopers* (doravante designada "PwC"), num estudo sobre *Smart Cities*[8], vem apresentar os seis desenvolvimentos tecnológicos e financeiros considerados críticos para impulsionar o crescimento das Smart Cities: i) as parcerias público-privadas; ii) o desenvolvimento de tecnologias emergentes (e.g., *blockchain, smart cars, IoT*); iii) a expansão da infraestrutura de TIC (e.g., evolução 4G, lançamento do 5G); iv) foco na cibersegurança (i.e., proteção da informações da cidade e dos dados dos cidadãos); v) *Cloud, edge and fog computing* (i.e., existe a necessidade de armazenamento em tempo real dos dados tendo em consideração o volume, a variedade e a velocidade dos mesmos); e vi) *Open data* e *Big Data Analytics*.

2.2. Os desafios para a privacidade

Os benefícios das *Smart Cities* são visíveis e fundamentais para ultrapassar os desafios dos tempos atuais, contudo, são inegáveis os desafios inerentes à implementação das mesmas. Para alcançarmos soluções inovadoras e inteligentes, onde conseguimos avaliar consumos e a qualidade da água, a qualidade do ar, a movimentação e tráfego na cidade, será necessária a recolha

7 Deloitte, Smart Cities, cit.

8 PricewaterhouseCoopers, Creating the Smart Cities of the future in Security and Privacy in your Smart City (Maio - 2019). Acessível em: <https://www.pwc.com/gx/en/sustainability/assets/creating-the-smart-cities-of-the-future.pdf> (consultado a 22 de novembro de 2020).

de grandes volumes e variedades de dados. A recolha massiva de dados e informações através de tecnologias de monitorização para a gestão das cidades inteligentes pode gerar insegurança para a privacidade dos cidadãos.

A recolha de dados através de tecnologias de *big data* e algoritmos e o respetivo tratamento dos mesmos são considerados o "coração" das *Smart Cities* ("*Data is the lifeblood of modern public policy*"[9]). Os dados recolhidos alimentam os estudos, as estatísticas, as análises e permitem o desenvolvimento de tecnologias inteligentes e adequadas a cada cidade consoante as suas especificidades. Contudo, o surgimento destas tecnologias vem acompanhado de vários desafios práticos e legais que devem ser devidamente estudados e acautelados.

De entre os riscos que podem advir para a privacidade dos cidadãos apontamos como merecedores de especial atenção os ciberataques, a gestão dos dados recolhidos, a não priorização da segurança e da privacidade no momento do desenvolvimento das novas tecnologias, ameaças físicas aos dispositivos, a falta de maturidade para os temas de privacidade e cibersegurança e ataques à integridade, confidencialidade e disponibilidade dos dados.

Se não forem tomadas medidas de forma a reduzir e mitigar riscos inerentes a ciberataques e ciberterrorismo, os dados podem ser comprometidos de tal forma de acarretem prejuízos incalculáveis[10].

Tendo em consideração os impactos e danos que podem advir de uma utilização indevida dos dados pessoais é necessário procurar um equilíbrio entre a privacidade dos cidadãos e a qualidade dos dados para efeitos das *Smart Cities*. Este equilíbrio pode ser alcançado através da implementação de várias medidas técnicas de organizativas adequadas para assegurar um nível de segurança adequado ao risco, como a pseudonimização e a anonimização (Cfr. Art. 32.º do RGPD).

Ainda que seja bastante visível o diminuto nível de maturidade da população relativamente ao tema da privacidade e da proteção de dados, a

9 Ranchordás, Klop, Data-driven Regulation, cit., 6.

10 Em 2017, a empresa dinamarquesa Maesk foi alvo do ataque cibernético mais devastador da história, o NotPetya, tendo o valor de dados totais sido estimado em $ 10 biliões. O NotPetya foi um malware que, através desta empresa, alcançou muitas outras empresas do mundo, tendo sido caracterizado como um ato de guerra. Sobre este tema consultar: <https://www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/> (Andy Greenberg, The Untold Story of NotPetya, the Most Devastating Cyberattack in History) .

regulação dos últimos anos nesta matéria permitiu o aumento do nível de awareness da população face a estes temas. Exemplo ilustrativo do aumento do conhecimento da população para estes temas, ocorreu em agosto de 2019 em Hong Kong, onde um conjunto de manifestantes, que procuravam lutar pela privacidade e pela proteção dos seus dados, derrubaram postes de luz "inteligentes" equipados de sensores e câmaras, por não se conformarem com a possível perda de privacidade[11].

2.3. A dicotomia entre os interesses públicos e privados

Os projetos de *Smart Cities*, como já referido anteriormente, assentam essencialmente na utilização de novas tecnologias para alcançar os seus objetivos, tecnologias estas que são providenciadas e desenvolvidas por empresas *Big Tech*. Estas empresas têm vindo a desenvolver soluções que permitem a mediação entre as cidades e os cidadãos e disponibilizam aos municípios não só os softwares, como também a implementação de sensores que recolhem dados dos cidadãos de forma massiva (*big data collection*) e que têm impacto na vida dos cidadãos.

Contudo, têm vindo a ser colocadas algumas questões relativamente aos interesses e valores destas empresas, em particular se estão ou não centrados nos cidadãos e nos valores e interesses públicos[12]. Estas questões têm vindo a ser levantadas por vários autores com fundamento na difícil conciliação entre os interesses e valores públicos e os interesses privados[13]. Isto porque as empresas *Big Tech* e as plataformas digitais estão imbuídas de valores privados centrados na maximização do lucro e no aumento da produtividade e eficiência que são substancialmente distintos dos valores e

11 Ranchordás, Sofia e Goanta, Catalina "The New City Regulators: Platform and Public Values in Smart and Sharing Cities", *University of Groningen Faculty of Law Research Paper Series*, No. 45/2019, (Out. - 2019).

12 Ranchordás e Goanta, The New City Regulators.

13 "While, for example, Huawei offers useful digital platforms for cities, it is also well-known that this company has been under investigation in different countries on suspicion of espionage and alleged trade-secret theft. This extreme example does not necessarily reflect the practices of other Big Tech companies, but it helps us illustrate the risks of a potential misalignment between public and private interests, the existence of hidden interests, and the lack of transparency of digital plat- forms." in Ranchordás, Goanta, The New City Regulators, cit.,8.

interesses públicos.

O potencial desequilíbrio entre os interesses públicos e privados levantam questões sobre a possível utilização indevida de dados pessoais de cidadãos e o desrespeito pela privacidade dos mesmos.

Os interesses e valores públicos que pretendem ser assegurados e alcançados com os projetos das *Smart Cities* são a qualidade de vida dos cidadãos e dos serviços públicos, a acessibilidade aos serviços públicos, a transparência, privacidade, a sustentabilidade e a igualdade do tratamento dos cidadãos.

No sentido de abordar este tema, Sofia Ranchordás e Catalina Goanta realizaram um estudo[14] onde procuraram compreender quais os potenciais conflitos de interesses e valores entre as empresas *Big Tech* e os interesses públicos e encontrar um equilíbrio entre os mesmos. Neste estudo foram apresentados alguns exemplos de potenciais interesses privados conflituantes, como é o caso da Huawei e do Airbnb. A Huawei, enquanto empresa que oferece soluções para cidades, foi alvo de investigações por suspeita de espionagem e roubo de segredos comerciais. Por outro lado, o Airbnb em Amesterdão causou uma crise imobiliária, colocando os direitos dos cidadãos e os valores públicos em causa. O caso do Airbnb é um exemplo de uma plataforma cujos interesses privados são conflituantes com interesses públicos, em particular refletido no enorme impacto causado nos moradores das cidades pelo aumento exponencial das casas em regime de alojamento local.

É inegável o desafio para estas empresas face à posição que assumem nestes projetos, na medida em que têm de se imbuir nos valores do bem público e gerar tecnologia dirigida para tal, afastando-se dos valores privados que lhes são inerentes. Contudo, ainda que os interesses possam ser distintos, tal não significa que os valores espelhados nas tecnologias e nos projetos desenvolvidos por empresas privadas não possam ser alinhados com os valores e interesses das empresas públicas.

Assim sendo, e na linha condutora do estudo elaborado por Sofia Ranchordás e Catalina Goanta[15], consideramos fundamental que no âmbito

14 Ranchordás e Goanta, The New City Regulators.
15 Ranchordás e Goanta, The New City Regulators.

dos projetos das Smart Cities sejam criadas normas legais que regulem esta colaboração de forma a garantir o acautelamento dos valores públicos e a boa-fé das empresas Big Tech nas negociações. A regulação desta relação tanto a nível legal como contratualmente é essencial para que se possa tirar proveito das tecnologias disponibilizadas pelas empresas garantindo que os valores públicos não são indevidamente prejudicados[16].

2.4. Privacy by design

O Regulamento Geral de Proteção de Dados, no seu art. 25.º, introduziu o princípio do *Privacy by Design*. Este princípio, desenvolvido na década de 90 pela Dra. Ann Cavoukian, procurou responder aos efeitos das tecnologias de informação e de comunicação na privacidade. O Princípio do *Privacy by Design* visa estabelecer de um padrão de atuação das organizações centrado na procura pela privacidade, que não é restringido ao mero cumprimento das normas legais existentes. O objetivo pretendido com a introdução deste princípio prende-se a antecipação da ponderação de todos os riscos que podem advir para os direitos, liberdades e garantias dos titulares dos dados, procurando mitigá-los à *priori*. Este princípio exige que as organizações incorporem a privacidade desde o design e a arquitetura dos projetos, sistemas ou práticas de negócio, tornando-se parte integrante dos mesmos.

O princípio do *Privacy by Design* é central no desenvolvimento dos projetos das *Smart Cities* tendo em consideração que consubstanciam projetos com recolha de grandes quantidades de dados pessoais, bem como através de tecnologias inovadoras, podendo implicar riscos para os direitos e liberdades dos titulares dos dados pessoais. O Princípio do *Privacy by Design*[17] implica

16 "In this context, we suggested a normative framework focusing on two points: departing from values shared by platforms and authorities, in order to shape a new kind of knowledge-service creation, namely local public-interest technology; and addressing the digital enforcement issue driven by the functional sovereignty role of platforms, by proposing a negotiated contractual system that seeks to balance platform values with public values." in Ranchordás e Goanta, The New City Regulators, cit., 32.

17 "*In essence, this means you have to integrate or 'bake in' data protection into your processing activities and business practices, from the design stage right through the lifecycle.*" in Information Commissioner's Office, Data protection by design and default. Acessível em: <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-by-design-and-

uma ponderação *ad initium* das questões relativas à proteção de dados, de forma a avaliar e prevenir potenciais impactos das tecnologias necessárias ao desenvolvimento dos projetos das *Smart Cities* na privacidade dos cidadãos.

O cumprimento deste princípio deve ser garantido pelas entidades públicas que desenvolvam os projetos através da implementação de procedimentos internos e medidas técnicas e organizativas, bem como por todas as entidades subcontratadas para o desenvolvimento dos mesmos. O cumprimento por parte dos subcontratantes deve ser garantido através da celebração de um acordo em matéria de proteção de dados, onde devem ser estabelecidas exigentes medidas de segurança da informação, nos termos e para os efeitos dos art. 28.º e 32.º do Regulamento Geral de Proteção de Dados.

3. Agregação de dados

A agregação de dados é uma técnica utilizada comummente para fins estatísticos através da qual os dados são apresentados de forma resumida. Esta técnica é atualmente utilizada pela grande maioria dos setores de atividade, desde o setor do *marketing* ao setor da saúde, na medida em que permite a análise de grandes quantidades de dados obtendo informações resumidas e de consulta rápida e eficiente. A agregação de dados é uma técnica essencial para potenciar o crescimento dos negócios e monetizar os dados na medida em que permite às organizações a obtenção de dados fundamentais para a compreensão do negócio e das tendências de mercado.

No âmbito de projetos de *Smart Cities* a utilização desta técnica permite uma otimização dos dados recolhidos, reduzindo o risco de inconformidade com a legislação em vigor em matéria de proteção de dados e construindo e reforçando a confiança dos cidadãos nos projetos.

Através da técnica de agregação, os dados são recolhidos e são tratados posteriormente de forma a que sejam criados dados resumidos para análise. O processo de agregação contempla três fases: Recolha, Tratamento e Apresentação. Em primeiro lugar, existe uma recolha de dados pessoais (e.g., através de fontes de IoT) e são armazenados em bases de dados. De seguida, os dados recolhidos são agregados através de funções estatísticas e, por último, os

default/> (consultado a 22 de novembro de 2020)

dados são apresentados de forma agregada num formato resumido.

A técnica da agregação é incluída na maioria das vezes nas técnicas de anonimização de dados visto que poderá ter um efeito semelhante à técnica da anonimização quando os elementos individuais que permitam a identificação de uma pessoa singular sejam substituídos por dados referentes a grupos de pessoas. A utilização desta técnica reduz significativamente o risco para a privacidade dos titulares dos dados. No entanto, ao longo do presente estudo vamos analisar se, de um ponto de vista legal, os dados agregados são subsumíveis ou não no conceito de dados pessoais.

Esta técnica é utilizada quando não é necessário manter os dados com identificações pessoais e a utilização de dados agregados é suficiente para alcançar as finalidades pretendidas.

Nas tabelas abaixo exemplificamos o método de anonimização por agregação[18]. Na primeira tabela encontram-se os dados após a recolha e na segunda são apresentados os dados agregados. Neste exemplo objeto de anonimização foram recolhidos os códigos postais e consumos de água por ano numa amostra de oito consumidores.

Tabela 1 - Conjunto de dados após a recolha.

PESSOA	CÓDIGO POSTAL	CONSUMO DE ÁGUA (M^3/ HAB.) POR ANO
Pessoa A	1750-123	64,5
Pessoa B	1700-456	44,9
Pessoa C	1700-654	48,7
Pessoa D	1750-231	70,5
Pessoa E	1070-111	38,9
Pessoa F	1070-232	41,2
Pessoa G	1570-400	81,3
Pessoa H	1570-223	78.9

Fonte: Elaboração própria

18 Os dados apresentados são fictícios, servindo apenas para a finalidade de compreensão da técnica de anonimização.

Tabela 2 - Conjunto de dados agregados.

CÓDIGO POSTAL	INTERVALO DE CONSUMO DE ÁGUA (M^3/HAB.) POR ANO
1750	64,5 - 70,5
1700	44,9 - 48,7
1070	38,9 - 41,2
1570	78,9 - 81,3

Fonte: Elaboração própria

No exemplo apresentado nas tabelas 1 e 2 foi utilizada uma amostra de 8 consumidores, o que aumenta em larga medida o risco de re-identificação dos titulares dos dados. Contudo, não deixa de ser importante notar que esta técnica reduz significativamente a utilidade e a qualidade dos dados, não podendo ser utilizado em todas as situações, sob pena dos dados deixarem de ser úteis.

4. Anonimização de Dados Pessoais

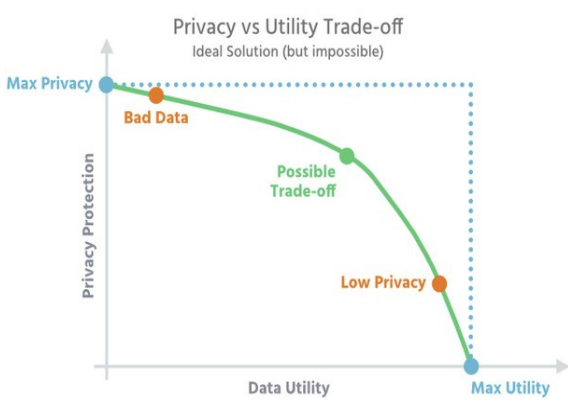
A técnica da anonimização surge nos últimos anos como uma ferramenta essencial para preservar a privacidade de conjuntos de dados que têm bastante valor e utilidade para as organizações. De forma a proteger a privacidade dos titulares dos dados, sempre que as organizações consigam manter a utilidade dos dados necessária para as finalidades estabelecidas, devem optar por esta técnica. Através da técnica da anonimização pretende-se balançar a utilidade dos dados pessoais e a privacidade dos mesmos de forma a que seja possível a transmissão de dados com redução da probabilidade de os mesmos serem associados a uma pessoa singular[19].

19 Parte da doutrina considera que este equilíbrio entre a utilidade dos dados e a anonimização é impossível de alcançar. Nas palavras de Paul Ohm "(…) research unearths a tension that shakes a foundational belief about data privacy: Data can be either useful or

A anonimização ideal é obtida através da maximização da privacidade e da utilidade dos dados (*trade-off*), o que é bastante difícil de alcançar, tendo em consideração que todos os modelos de anonimização utilizados padecem de limitações e que, na maioria dos casos, existe uma necessidade de interligação de registos entre bases de dados distintas.

O gráfico abaixo pretende facilitar a compreensão da relação entre a privacidade e a utilidade dos dados pessoais, sendo possível concluir que, à medida que ganhamos privacidade relativamente a um conjunto de dados pessoais, a utilidade dos mesmos diminuí. A diminuição da privacidade consubstancia-se essencialmente na possibilidade de terceiros re-identificarem pessoas singulares. O aumento da utilidade dos dados ocorre quando se aumenta a quantidade de informações sobre os indivíduos, obtendo informações mais completas[20].

Figura 2 - Privacidade vs. Utilidade dos dados pessoais



Fonte: SARTOR, Nicolas. Data Anonymisation Software – Differences Between Static and Interactive Anonymisation. Disponível em: <https://www.datasciencecentral.com/profiles/blogs/data-anonymisation-software-differences-between-static-and>

No entanto, é fundamental ter em consideração que a figura é meramente ilustrativa e que a comparação destas duas variáveis (i.e., utilidade

perfectly anonymous but never both." In Ohm, Paul, Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization, *57 UCLA Law Review* 1701 (2010), cit., 1704.

20 Tiancheng Li, Ninghui Li, On the Tradeoff Between Privacy and Utility in *Data Publishing*, Department of Computer Science Purdue University (2009).

e privacidade dos dados) não representa na realidade um ganho e uma perda proporcionais, variando de acordo com vários fatores incluindo a utilização de dados individuais ou dados agregados[21].

4.1. Anonimização e pseudonimização

A anonimização e a pseudonimização (ou utilização de pseudónimos) são suas técnicas que, embora sejam muitas vezes confundidas, representam duas realidades distintas e com diferentes impactos para a proteção de dados.

O processo de anonimização permite a conversão irreversível de dados pessoais em dados não identificáveis. Por sua vez, a pseudonimização é definida nos termos do n.º 5 do art. 4.º, do Regulamento Geral de Proteção de Dados como um *"(...) [t]ratamento de dados pessoais de forma a que deixem de poder ser atribuídos a um titular de dados específico sem recorrer a informações suplementares, desde que essas informações suplementares sejam mantidas separadamente e sujeitas a medidas técnicas e organizativas para assegurar que os dados pessoais não possam ser atribuídos a uma pessoa singular identificada ou identificável*". A pseudonimização permite que o Responsável pelo Tratamento[22] associe os dados aos titulares dos dados pessoais através de um processo em que se substituí todos os identificadores pessoais por pseudónimos, que podem ser códigos gerados artificialmente para mascarar os dados originais.

A grande distinção entre ambos consiste no carácter definitivo da anonimização e na possibilidade de reversibilidade da pseudonimização. Consequentemente, a legislação em matéria de proteção de dados é aplicável à pseudonimização, na medida em que permite a identificação dos titulares dos dados, ainda que indiretamente. Ao invés, a anonimização, constituindo um tratamento irreversível, é excluída do âmbito de aplicação da legislação relativa à proteção de dados.

Contudo, cumpre notar que a pseudonimização não deixa ter uma enorme importância enquanto medida técnica, nos termos do art. 32.º do RGPD,

21 Tiancheng Li e Ninghui Li, *On the Tradeoff*.

22 O conceito de Responsável pelo Tratamento deve ser entendido nos termos e para os efeitos do n.º 7 do art.4.ºdo Regulamento Geral de Proteção de Dados.

para efeitos de segurança de informação. A utilização desta técnica reduz a capacidade de ligação de um conjunto de dados à identidade dos seus titulares e, para obtenção de uma pseudonimização com um caráter mais "forte", são utilizados mecanismos de atribuição de códigos aleatórios.

4.2. Anonimização técnica

A anonimização é uma técnica através da qual se pretende obter a conversão de dados identificáveis em dados não identificáveis.

A anonimização encontra-se definida em normas internacionais, como é o caso da ISO 29100:2011[23] onde é defina como "*process by which personally identifiable information (PII) is irreversibly altered in such a way that a PII principal can no longer be identified directly or indirectly, either by the PII controller alone or in collaboration with any other party*".

O processo de anonimização visa tornar os dados anónimos, no entanto existem diversas técnicas que podem ser utilizadas para o efeito. Estas técnicas, ainda que com o mesmo objetivo, apresentam níveis de re-identificação distintos variando de acordo com o *data set* utilizado.

As diferentes técnicas de anonimização devem ser utilizadas consoante a situação em causa, na medida em que técnica tem especificidades que podem ser benéficas para diferentes tipos de situações.

Existem duas abordagens distintas para alcançar a anonimização dos dados pessoais, em particular, a Aleatorização e a Generalização, que passamos abaixo a detalhar.

4.2.1. Aleatorização (Randomization)

A técnica da Aleatorização (*Randomization*) visa a alteração da veracidade dos dados pessoais de forma que não seja possível a ligação dos dados às pessoas em causa. Esta técnica baseia-se na introdução de um fator de incerteza nos dados, contribuindo para a diminuição da associação entre os dados e as pessoas singulares.

23 ISO/IEC 29100:2011, Information technology — Security techniques — Privacy framework

As principais técnicas de anonimização por aleatorização são a introdução de ruído (*Noise Addition*), a permutação (*Shuffling*) e a privacidade diferencial (*Differential Privacy*). Para garantir a total irreversibilidade dos dados, podem ser utilizadas várias técnicas em simultâneo.

A técnica da introdução de ruído é uma das técnicas mais utilizada e consiste na introdução de ruído nos dados de forma a conferir confidencialidade aos mesmos. Contudo, tem-se vindo a constatar que com a tentativa de alcançar uma maior confidencialidade dos dados e um menor risco de re-indetificação, esta técnica tem vindo a perder algumas propriedades estatísticas.

4.2.2. Generalização

A generalização é uma técnica de anonimização de dados que consiste em generalizar ou diluir os dados pessoais através da alteração da escala ou ordem de grandeza[24].

As principais técnicas de anonimização por generalização são a agregação[25] e kanonimato e a L-diversidade/t-proximidade.

Uma das técnicas mais utilizadas, e com maior importância para o estudo em causa, é a técnica da agregação. As técnicas de agregação e k-anonimato "(…) visam impedir que um titular dos dados seja selecionado através do agrupamento com, pelo menos, outras k pessoas"[26]. Esta técnica permite que os titulares dos dados não sejam identificados na medida em que os dados pessoais são partilhados por vários utilizadores (k utilizadores). Os dados são generalizados de forma a que vários titulares partilhem dos mesmos dados.

5. Anonimização Legal

24 GT 29, Parecer 05/2014 do Grupo de trabalho do artigo 29.º sobre técnicas de anonimização, adotado em 10 abril de 2014.

25 A técnica da agregação encontra-se desenvolvida no capítulo "D. Agregação de Dados".

26 GT 29, Parecer 05/2014.

Como já foi referido anteriormente, a anonimização é uma técnica que é aplicada a dados pessoais com o objetivo de evitar de forma irreversível a identificação do titular dos dados e é antecedida de um processo de recolha de dados que carece do cumprimento da legislação em vigor em matéria de proteção de dados pessoais.

O Regulamento Geral de Proteção de Dados entre os princípios relativos ao tratamento de dados pessoais, estabelece o princípio da limitação das finalidades, nos termos da alínea b) do n.º 1 do art. 5.º, nos termos do qual os dados pessoais devem ser recolhidos para finalidades determinadas, explícitas e legítimas, não podendo ser tratados posteriormente de forma incompatível com essas finalidades. Este princípio consagra ainda que não é considerado incompatível com as finalidades iniciais, o tratamento posterior para fins de arquivo de interesse público, ou para fins de investigação científica ou histórica ou para fins estatísticos.

Ora, através da técnica da anonimização é possível a utilização dos dados para finalidades distintas, sem a limitação que nos é exigida pelo princípio da limitação das finalidades.

A Diretiva 95/46/CE, no seu considerando 26, excluía do seu âmbito de aplicação os dados objeto de anonimização, referindo expressamente que *"(...) os princípios da protecção não se aplicam a dados tornados anónimos de modo tal que a pessoa já não possa ser identificável (...)*".

Com a entrada em vigor do Regulamento Geral de Proteção de Dados que veio revogar a Diretiva 95/46/CE, o legislador manteve a exclusão da aplicação da legislação relativa a proteção de dados pessoais aos dados anonimizados. Nos termos do considerando 26 do mesmo, é estabelecido que *"(...) [o]s princípios da proteção de dados não deverão, pois, aplicar-se às informações anónimas, ou seja, às informações que não digam respeito a uma pessoa singular identificada ou identificável nem a dados pessoais tornados de tal modo anónimos que o seu titular não seja ou já não possa ser identificado. O presente regulamento não diz, por isso, respeito ao tratamento dessas informações anónimas, inclusive para fins estatísticos ou de investigação*".

O considerando 26 do RGPD oferece-nos uma definição de anonimização detalhando que os dados pessoais são considerados anónimos

quando não dizem respeito a uma pessoa identificada ou identificável. Ora, consequentemente, quando uma informação "diz respeito a uma pessoa singular identificada ou identificável", esta informação é qualificável como "dado pessoal" nos termos e para os efeitos da alínea a) do n.º 1 do art. 4.º do RGPD. Assim sendo, parece claro que para determinarmos se um conjunto de dados se encontra anonimizado é necessário avaliar se a informação permite identificar ou torna identificável os titulares dos dados.

Adicionalmente, o considerando 26 do RGPD avança ainda um critério para determinar se uma pessoa singular é ou não identificável. O critério escolhido pelo legislador para aferição da identificabilidade de pessoas singulares foi o critério da razoabilidade. Assim sendo, deverão ser tomados em consideração "(…) todos os meios suscetíveis de ser razoavelmente utilizados (…)".

Para o entendimento cabal do conceito que nos é oferecido pelo Regulamento Geral de Proteção de Dados de anonimização é crucial compreendermos o significado de dado pessoal e dos elementos que o compõem, nos termos do n.º 1 do art. 4.º do mesmo.

5.1. Definição de dados pessoais

O estudo do conceito de Dados Pessoais é fundamental para compreendermos o conceito de anonimização.

O conceito de dados pessoais consta do n.º 1 do art. 4.º do RGPD consubstanciando-se em toda a "informação relativa a uma pessoa singular identificada ou identificável («titular dos dados»)". O referido art. acrescenta ainda que *"(...) é considerada identificável uma pessoa singular que possa ser identificada, direta ou indiretamente, em especial por referência a um identificador, como por exemplo um nome, um número de identificação, dados de localização, identificadores por via eletrónica ou a um ou mais elementos específicos da identidade física, fisiológica, genética, mental, económica, cultural ou social dessa pessoa singular*".

A definição de dados pessoais que nos é concedida pelo Regulamento Geral de Proteção de Dados é composta essencialmente por 5 elementos, em particular: i) informação; ii) relativa a; iii) uma pessoa singular; iv) identificada;

e v) identificável.

Para um total entendimento do conceito de dados pessoais é necessária a compreensão detalhada dos elementos que o constituem.

a) Informação

A expressão "informação" denota a intenção do legislador em criar um conceito amplo de dados pessoais. O Grupo de Trabalho do Artigo 29[27], relativamente a este conceito, incluí todas as informações objetivas ou factuais (e.g., determinado parâmetro num exame médico, nacionalidade de uma pessoa) e as informações subjetivas (e.g., opiniões, avaliações, testes psicotécnicos) sobre determinada pessoa.

A informação tem relevância para o Direito da Proteção de Dados independentemente do suporte ou formato em que for recolhida e armazenada, incluindo o formato alfabético, fotográfico, numérico, gráfico, suporte papel, código binário, CD ou cassete[28].

Por outro lado, não com menor relevância, têm-se vindo a discutir se a informação no contexto do n.º 1 do art. 4º, tem de ser verdadeira. Alguns autores têm vindo a defender que a informação não necessita de ser verdadeira ou comprovada, com base no argumento de que o próprio Regulamento Geral de Proteção de Dados, no seu art. 16.º, confere ao titular dos dados o direito de solicitar a retificação dos dados que considere inexatos. Ora, o próprio RGPD parte do pressuposto que os dados podem não se encontrar exatos e verdadeiros, conferindo o direito à retificação, não excluindo tais dados do seu âmbito de aplicação.

No que concerne ao conteúdo da informação, o conceito de dados pessoais abrange os mais variados aspetos relativamente ao titular dos dados pessoais, nomeadamente físicos, mentais, familiares ou sociais. As informações sobre os titulares incluem dados identificativos (e.g., nome, número de identificação, data de nascimento), dados de localização, características físicas (e.g., peso, altura, cor da pele, cabelo, olhos) e identificadores por via eletrónica (Cfr. art.

27 GT 29, Parecer 4/2007 do Grupo de trabalho do art. 29.º sobre o conceito de dados pessoais, adotado em 20 de junho.

28 *Ibidem*.

4º, n.º 1 do RGPD). O conceito de informação abrange ainda todos os dados que revelem a origem racial ou étnica, opiniões políticas ou filosóficas, filiação sindical, dados genéticos, dados biométricos que identifiquem uma pessoa de forma inequívoca, dados relativos à saúde ou dados relativos à vida sexual (Cfr. art. 9.º do RGPD).

b) Relativa a

A expressão "relativa a" vem restringir a informação à pessoa a que respeita. Para que a informação seja considerada um dado pessoal tem de estar relacionada com uma pessoa, tem de ser informação sobre uma pessoa.

Nas palavras de A. Barreto Menezes Cordeiro *"[e]sta intrínseca relação entre a informação e um sujeito excluí o campo de aplicação do RGPD toda a informação concernente a realidades jurídicas não subjetiváveis*".

Algumas informações são facilmente associadas a pessoas singulares, contudo existem informações que apresentam uma maior dificuldade na determinação da relação com pessoas singulares. Quando as informações são referentes a objetos ou constituem dados factuais é necessária uma avaliação casuística para a determinação de uma possível relação com uma pessoa singular. Um exemplo elucidativo que nos é apresentado pelo Grupo de Trabalho do Artigo 29.º relativamente ao conceito de dados pessoais prende-se com o registo de um carro numa oficina que contempla variadas informações sobre o carro (e.g., os quilómetros, as revisões, as condições do material). O Grupo de Trabalho do Artigo 29.º vem-nos dizer que, analisados os dados individualmente, poderiam não constituir dados pessoais, contudo, relacionado com o proprietário para efeitos de faturação, constitui informação "relativa ao" proprietário[29].

O Grupo de Trabalho do Artigo 29.º, a este respeito, vem sintetizar três elementos alternativos que permitem considerar uma informação "relativa a" uma determinada pessoa: os elementos de conteúdo, de finalidade e de resultado[30]. O elemento de conteúdo representa toda a informação sobre uma pessoa singular (e.g., os exames médicos que são realizados a um doente,

29 GT 29, Parecer 4/2007.
30 *Ibidem*.

logo essa informação respeita ao doente). No elemento da finalidade, os dados utilizados têm como finalidade "avaliar, tratar de determinada forma ou influenciar o estatuto ou comportamento de uma pessoa"[31]. No que respeita ao elemento de resultado, estão em causa todas as informações que são relativas a uma pessoa porque o tratamento das mesmas pode ter impactos nos direitos e interesses da pessoa em causa.

c) Uma pessoa singular

O Regulamento Geral de Proteção de Dados restringe a aplicação dos princípios de proteção de dados a informações relativas a pessoas singulares independentemente da sua nacionalidade ou do seu local de residência (Cfr. Considerandos 14 e 26 do RGPD). No considerando 14 do RGPD é excluída a proteção conferida pelo mesmo a pessoas coletivas ficando ainda de fora "*(...) demais realidades jurídicas não subjetiváveis, como coisas e os animais*"[32].

No que concerne à exclusão dos objetos cumpre alertar que poderão existir informações relativas a objetos que possam constituir dados pessoais de determinada pessoa singular. Pegando novamente no exemplo do registo de um veículo numa oficina, onde constam as informações sobre o veículo, é evidente que as informações relativas ao veículo em si não constituem dados pessoais, no entanto, se relacionadas com uma pessoa singular podem constituir dados pessoais.

Ainda que a legislação em matéria de proteção de dados exclua do seu âmbito de aplicação a proteção de pessoas coletivas, importa ter em consideração que, em determinados casos, poderá ser aplicável a informações relativas a empresas e/ou pessoas coletivas. Quando a informação relativa a empresas e/ou pessoas coletivas também seja relativa a pessoas singulares (quando verificado um dos elementos de "conteúdo", "finalidade" ou "resultado"), também deverá ser qualificada como dado pessoal e consequentemente devem ser aplicadas as regras relativas à proteção de dados.

O considerando 27 do RGPD esclarece ainda que o Regulamento "(…)

31 GT 29, Parecer 4/2007.

32 Cordeiro, A. Barreto Menezes, Dados pessoais: conceito, extensão e limites, Revista de Direito e Tecnologia, Vol.1 (2019), n.º 2 (2018), p. 297-321, 10. Acessível em: <https://blook.pt/publications/fulltext/e38a9928dbce/>

não se aplica aos dados pessoais de pessoas falecidas". No entanto, a Lei n.º 58/2019, de 8 de agosto, que assegura a execução do Regulamento Geral de Proteção de Dados, no ordenamento jurídico português, no seu art. 17.º, vem alargar o âmbito de aplicação do RGPD, protegendo ainda os dados de pessoais de pessoas falecidas quando se tratem de categorias especiais de dados pessoais (art. 9.º do RGPD), quando se reportem à intimidade da vida privada, à imagem ou aos dados relativos às comunicações. Os direitos consagrados no Regulamento Geral de Proteção de Dados podem ser exercidos por quem tenha sido designado para o efeito pela pessoa falecida ou, na sua falta, pelos seus herdeiros (Cfr. art. 17.º, n.º 2 do RGPD). No entanto, se os dados disserem respeita ao falecido e a outro titular vivo, a informação deixa de poder ser livremente utilizada.

d) Conceito de pessoa identificada

O Regulamento Geral de Proteção de Dados define dados pessoais como a informação relativa a uma pessoa singular identificada ou identificável (Cfr. art. 4.º, n.º 1 do RGPD). O Regulamento, acrescenta ainda que "é considerada identificável uma pessoa singular que possa ser identificada, direta ou indiretamente, em especial por referência a um identificador, como por exemplo um nome, um número de identificação, dados de localização, identificadores por via eletrónica ou a um ou mais elementos específicos da identidade física, fisiológica, genética, mental, económica, cultural ou social dessa pessoa singular".

O legislador utilizou a expressão "direta" para caracterizar as situações em que uma pessoa singular é identificada sem necessidade de recorrer de recorrer a dados adicionais. O Tribunal de Justiça da União Europeia, no Acórdão Breyer, vem esclarecer relativamente ao tema dos IPs dinâmicos dos utilizadores que "(…) deve-se, antes de mais, salientar que é dado assente que um endereço IP dinâmico não constitui uma informação relativa a uma «pessoa singular identificada», na medida em que esse endereço não revela diretamente a identidade da pessoa singular proprietária do computador a partir do qual se efetua a consulta de um sítio Internet, nem a de outra pessoa que possa utilizar esse computador".

Schwartz, a este respeito, vem referir que "*[a]mong EU member states that have traditionally taken a leading role in information privacy law, a person falls in the "identified" category if a party can use information relating to her to determine her specific identity*"[33].

Uma pessoa considera-se identificada quando a informação em causa respeita diretamente a essa pessoa e seja suficiente para a identificar inequivocamente, sem que sejam necessárias informações adicionais[34]. Exemplos de dados que respeitam a uma pessoa singular identificada são o nome completo de uma pessoa, o cartão de cidadão, a sua impressão digital, o NIF.

e) Conceito de pessoa identificável

O conceito de pessoa identificável é um conceito-chave para a verdadeira compreensão do significado de dados pessoais e de anonimização. Contudo, este conceito é complexo e tem vindo a ser objeto de diversas e distintas interpretações.

Para que uma pessoa seja considerada identificável é necessário que seja possível identificar o titular dos dados com informações adicionais sobre o mesmo. Ou seja, ainda que determinado dado não permita identificar o titular dos dados, conjugada com outra informação é possível a identificação do titular.

Nos termos do considerado 26, o Regulamento refere ainda que, "*[p]ara determinar se uma pessoa singular é identificável, importa considerar todos os meios suscetíveis de ser razoavelmente utilizados, tais como a seleção, quer pelo responsável pelo tratamento quer por outra pessoa, para identificar direta ou indiretamente a pessoa singular. Para determinar se há uma probabilidade razoável de os meios serem utilizados para identificar a pessoa singular, importa considerar todos os fatores objetivos, como os custos e o tempo necessário para a identificação, tendo em conta a tecnologia disponível à data do tratamento dos dados e a evolução tecnológica*"[35].

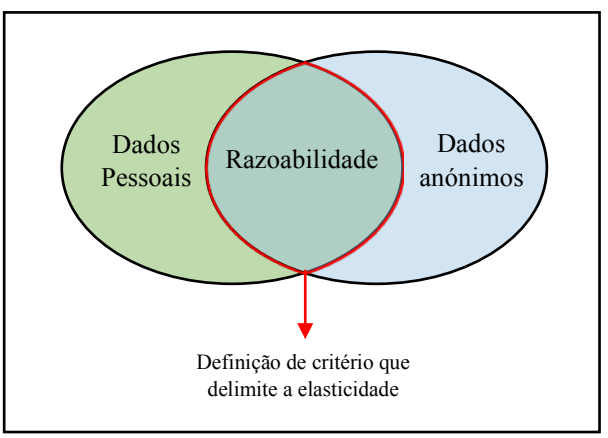
33 Schwartz, Paul M. e Solove, Daniel J., Reconciling Personal Information in the United States and European Union, Vol. 102:877, *California Law Review,* (2014), cit., 882.

34 Cordeiro, A. Barreto Menezes, Dados pessoais, cit., 14 ss.

35 TJUE 19-out.-2016, proc. C-582/14 (Acórdão Breyer).

O Regulamento Geral de Proteção de Dados apresenta-nos um conceito amplo e indeterminado de dados pessoais o que torna mais complexa a distinção entre dados pessoais e dados anonimizados. Assim sendo, implica a necessidade de definição de um critério que permita a distinção entre dados pessoais e dados anonimizados sob pena da fronteira entre ambos ser transponível. Este critério deverá delimitar da elasticidade do conceito de dados pessoais, em particular o termo identificável.

Figura 3 – Dados Pessoais vs Dados Anónimos



Fonte: Elaboração própria

O Regulamento Geral de Proteção de Dados apresenta-nos como critério para aferição do conceito de "identificável" o critério da razoabilidade. A interpretação do conceito de "identificável" será interpretado casuisticamente pelo intérprete, consoante os dados em causa e com os meios suscetíveis de serem razoavelmente utilizados para identificar direta ou indiretamente a pessoa singular. Nas palavras de A. Barreto Menezes Cordeiro *"[a] assunção de um critério de razoabilidade é um reflexo da impossibilidade fáctica de garantir a anonimidade absoluta dos dados recolhidos*".

O critério da razoabilidade que nos é oferecido pelo RGPD visa limitar a elasticidade do conceito de dados pessoais restringindo a qualificação do conceito de dados pessoais a um esforço razoável em obtê-los. Este critério

deve ser interpretado de acordo com dois tipos de fatores: i) fatores objetivos; e ii) fatores subjetivos. Os fatores objetivos, encontram-se elencados no considerando 26 do RGPD e consistem nos seguintes: custos, tempo necessário para a identificação, a tecnologia disponível à data do tratamento de dados e a evolução tecnológica. Estes fatores implicam uma análise dinâmica e que varia consoante o momento do tratamento de dados pessoais. Os fatores subjetivos apresentados pelo legislador relativamente ao critério da razoabilidade prendem-se com os meios suscetíveis de serem razoavelmente utilizados pelo responsável pelo tratamento ou por outra pessoa. Estes fatores respeitam à capacidade individual de engenharia reversa de quem procede ao tratamento dos respetivos dados pessoais.

O legislador procurou apresentar um critério tecnologicamente neutro, não apontando diretamente para nenhuma tecnologia concreta[36]. O objetivo desta escolha foi o desenho de uma norma que permite o acompanhamento constante do desenvolvimento tecnológico, tendo em consideração que ao longo deste desenvolvimento surgirão cada vez mais tecnologias que tornarão mais provável a re-identificação de dados anonimizados. Assim sendo, o conceito de dados pessoais deve também ser avaliado no momento concreto em que se pretende efetuar um tratamento de forma a ser ponderado o estado tecnológico do momento.

No que respeita aos meios suscetíveis de serem razoavelmente utilizados pelo responsável pelo tratamento ou por outra pessoa existem dois problemas que são apresentados pela doutrina e jurisprudência: i) a possibilidade de serem utilizados meios ilícitos para a identificação do titular dos dados; ii) a relevância da informação obtida através de meios suscetíveis de serem razoavelmente utilizados por outras pessoas.

No que concerne à licitude dos meios, o Tribunal de Justiça da União Europeia (doravante designado "TJUE"), no Acórdão Breyer o advogado-geral considerou que *"(...) assim não será se a identificação da pessoa em causa for proibida por lei ou inexequível, por exemplo devido ao facto de implicar um esforço desmedido em termos de tempo, de custo e de mão de obra, de*

36 "*A fim de se evitar o sério risco sério de ser contornada a proteção das pessoas singulares, esta deverá ser neutra em termos tecnológicos e deverá ser independente das técnicas utilizadas*", considerando 15 do RGPD.

modo que o risco de uma identificação parece na realidade insignificante"[37]. Em suma, o presente acórdão vem considerar que as informações obtidas por meios ilícitos não devem ser tomados em consideração para efeitos do conceito de "identificável" sob pena de implicar um esforço desmedido.

Contudo, a posição do TJUE tem vindo a ser bastante criticada pela doutrina. Os argumentos apresentados pela doutrina em sentido contrário ao acórdão Breyer centramse essencialmente nas diferenças legislativas que existem nos vários ordenamentos jurídicos podendo em determinados países ser considerado um meio lícito e noutros ilícito, bem como a frequência da ocorrência de ataques informáticos[38]. Nesta sequência, A. Barreto Menezes Cordeiro afirma que "*(...) sendo o critério último o da razoabilidade, importa atender a todas as condutas ilícitas que possam razoavelmente contar com elas*"[39]. O autor baseia a sua afirmação em três elementos interpretativos, o elemento literal, através do qual refere que o critério de razoabilidade apresentado no considerando 26 não excluí ilicitudes, o elemento teleológico onde é realçado o objetivo do Regulamento Geral de Proteção de Dados em defender a devassa da vida privada no seu todo, e o elemento sistemático que permite a conclusão de que o direito da proteção de dados foi criado tendo em consideração que os dados dos titulares podem ser obtidos ilicitamente[40].

No que respeita à segunda problemática abordada e debatida pela doutrina é discutida a relevância da informação obtida através de meios suscetíveis de serem razoavelmente utilizados por outras pessoas. Quanto a esta problemática a doutrina tem vindo a dividir-se essencialmente, ainda que com algumas variações, em duas teorias: a teoria relativa e a teoria absoluta.

Por um lado, a teoria relativa vem considerar que para a determinação do conceito de "pessoa singular identificável" devem ser apenas tidos em consideração os meios e conhecimentos detidos pelo Responsável pelo Tratamento. Por outro lado, os defensores a teoria absoluta consideram que devem não só ser tomados em consideração os meios e conhecimentos detidos

37 TJUE 19-out.-2016, proc. C-582/14 (Acórdão Breyer).

38 Cordeiro, A. Barreto Menezes, Direito da Proteção de Dados à Luz do RGPD e da Lei n.º 58/2018, Almedina, Coimbra (2020) 123.

39 Cordeiro, A. Barreto Menezes, Direito da Proteção, 123.

40 *Ibidem*.

pelo Responsável pelo Tratamento como por terceiros.

A teoria relativa é defendida pela maioria da doutrina e jurisprudência alemã[41]. Um dos principais argumentos dos defensores da teoria relativa prende-se com a impossibilidade de um anonimato absoluto e irreversível. Neste sentido, Paul Ohm conclui que "*Computer scientists have recently undermined our faith in the privacy protecting power of anonymization, (...). These scientists have demonstrated that they can often "reidentify" or "deanonymize" individuals hidden in anonymized data with astonishing ease. By understanding this research, we realize we have made a mistake, labored beneath a fundamental misunderstanding, which has assured us much less privacy than we have assumed. This mistake pervades nearly every information privacy law, regulation, and debate, yet regulators and legal scholars have paid it scant attention*".

O Regulamento Geral de Proteção de Dados, no considerando 4, reconhece que o direito à proteção de dados não é absoluto e enaltece a dimensão económica do tratamento de dados pessoais. O direito à proteção de dados, não constitui um direito absoluto, e deve ser analisado de acordo com a sua função na sociedade e ser equilibrado com outros direitos fundamentais. Adicionalmente, a Comissão Europeia elencou como um dos grandes objetivos do Regulamento Geral de Proteção de Dados, para além da devolução do controlo aos titulares dos dados, o desenvolvimento da economia digital.

O enaltecimento da dimensão económica do tratamento dos dados pessoais é outro argumento central e fundamental para a compreensão da teoria relativa. O alargamento excessivo do conceito de dados pessoais vai pôr em causa a dimensão económica do tratamento dos dados e os interesses económicos da União Europeia e dos seus estados-membros face ao processo de globalização, limitando as possibilidades de utilização dos mesmos.

Adicionalmente, caso aplicássemos a teoria objetiva pura, todos os dados seriam considerados dados pessoais, na medida em dificilmente se alcança a inexistência de risco de re-identificação, o que, inevitavelmente, vai ter impacto tanto na dimensão económica do direito.

41 Cordeiro, A. Barreto Menezes, Direito da Proteção, 127.

Por outro lado, a doutrina defensora da teoria relativa argumenta ainda que:

i) A atribuição da relevância aos meios e conhecimentos detidos por terceiros impossibilitaria que o Responsável pelo Tratamento conhecesse se está ou não a respeitar a legalidade, tendo em consideração que esta poderá respeitar a um responsável sediado noutro país;

ii) Tornaria o processo de supervisão por parte das entidades responsáveis excessivamente pesado ou praticamente impossível;

iii) Consubstanciaria uma colisão de interesses, entre os interesses pessoais (i.e., proteção dos seus dados pessoais) e os interesses coletivos (i.e., livre iniciativa económica);

iv) Colocaria em causa a possibilidade de realização de estudos estatísticos.

No considerando 26 do RGPD o legislador faz referência a "*(...) todos os meios suscetíveis de ser razoavelmente utilizados, (...) quer pelo responsável pelo tratamento quer por outra pessoa*". Neste considerando é visível a referência do legislador face aos meios utilizados por terceiros. No entanto, de acordo com os defensores da teoria relativa esta expressão não conduz a uma teoria absoluta, na medida em que não se refere a todos os meios disponíveis por terceiros, mas apenas aqueles que possam ser razoavelmente utilizados. E os meios que podem ser razoavelmente utilizados por terceiros também devem ser levados em consideração pelo Responsável pelo Tratamento no âmbito da teoria relativa.

Face aos argumentos apresentados pela doutrina defensora da teoria relativa, os defensores da teoria objetiva consideram que sendo o critério aplicável o da razoabilidade, esta teoria nunca poderia ser objetiva na aceção pura no termo[42].

O Tribunal de Justiça da União Europeia, no acórdão Breyer, pronunciou-se acerca desta questão considerando que *"(...) um endereço IP dinâmico registado por um prestador de serviços de meios de comunicação em linha aquando da consulta por uma pessoa de um sítio Internet que esse prestador disponibiliza ao público constitui, relativamente a esse prestador,*

42 Cordeiro, A. Barreto Menezes, Direito da Proteção, 128.

um dado pessoal na aceção dessa disposição, quando este disponha de meios legais que lhe permitam identificar a pessoa em causa graças às informações suplementares que o fornecedor de acesso à Internet dessa pessoa dispõe"[43]. Face à presente decisão do TJUE, A. Barreto Menezes Cordeiro vem referir que *"[a] posição sufragada pelo TJUE não é fácil de catalogar, na medida em que rejeita a teoria relativa, mas, ao mesmo tempo, rejeita a teoria objetiva no seu estado mais puro*"[44].

6. Utilização dos dados agregados no âmbito dos projetos de Smart Cities

6.1. Contexto

Este estudo é centrado na privacidade dos dados da comunidade na medida em que consideramos que os pilares essenciais da arquitetura de uma cidade inteligente são a transparência, confiança dos cidadãos e respeito pela privacidade dos mesmos. Apenas através da criação de um projeto totalmente transparente perante os cidadãos se consegue alcançar o sucesso do projeto. Por outro lado, é fundamental que desde a origem do projeto seja proactivamente tomada em consideração a privacidade e sejam desenhadas medidas técnicas e organizativas para garantir o total respeito pela legislação em matéria de proteção de dados, bem como prevenir e antecipar riscos (Princípio do *Privacy by Design*). O reconhecimento do valor e dos benefícios da adoção de medidas fortes ao nível da privacidade é essencial para garantir o respeito pelos direitos, liberdades e garantias dos titulares dos dados.

Através do presente estudo, pretendemos debruçar-nos sobre uma solução que por um lado garanta o respeito pelos direitos fundamentais dos cidadãos, em particular, o direito à privacidade e, por outro, permita a utilização dos dados dos mesmos para efeitos de desenvolvimento de cidades inteligentes. No que respeita ao desenvolvimento das cidades inteligentes, vamos centrar o nosso estudo no aproveitamento e otimização dos dados que já foram recolhidos para determinadas finalidades de forma a ser possível a utilização para outras finalidades, que não careçam da identificação dos titulares dos dados. Estas

43 TJUE 19-out.-2016, proc. C-582/14 (Acórdão Breyer).

44 Cordeiro, A. Barreto Menezes, Direito da Proteção, 129.

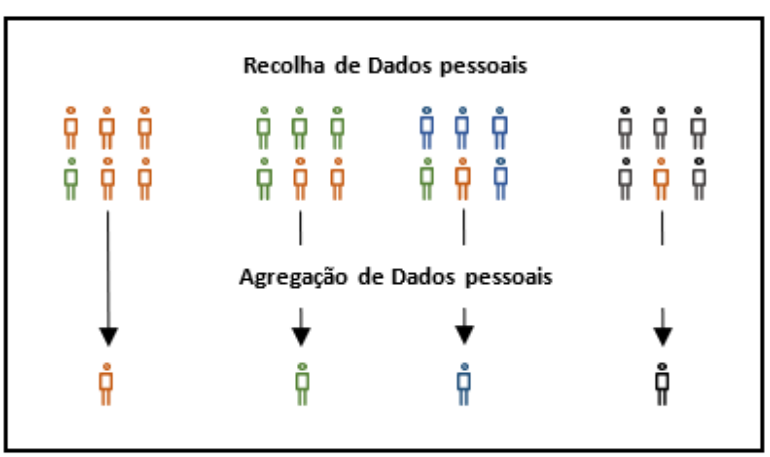
finalidades, no âmbito dos projetos das *Smart Cities*, são a caracterização da população e a compreensão das tendências para a prestação de um melhor serviço público. Analisadas as características e tendências dos munícipes, é possível a adequação de medidas, políticas, serviços e produtos, bem como a criação de novos produtos e serviços de forma a proporcionar uma maior qualidade de vida aos cidadãos.

Para tal, vamos estudar a técnica da agregação, de um ponto de vista legal, de forma a compreender se os dados agregados são dados anónimos ou se são subsumíveis ao conceito de dados pessoais constante do RGPD. Isto porque, se os dados agregados forem considerados dados pessoais, ficam sujeitos às exigências do RGPD e, por consequência, tem de ser respeitado o princípio da limitação das finalidades, que não permite o tratamento posterior de forma incompatível com as finalidades que motivaram a recolha (art. 5.º, n.º 1, alínea b) do RGPD).

Para compreendermos os tratamentos de dados pessoais em causa, consideramos essencial a divisão do projeto em duas fases:

a) A fase da recolha e da agregação dos dados (doravante designada de "Fase 1")

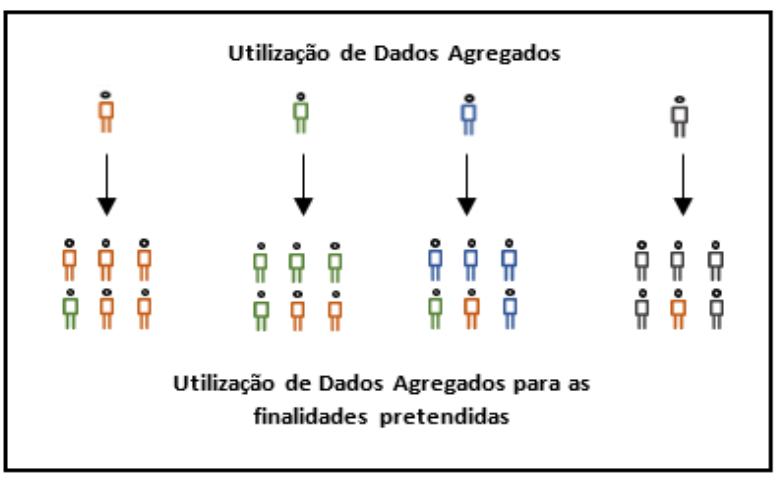
Figura 4 – Recolha e Agregação de Dados (Fase 1)



Fonte: Elaboração própria

b) A fase de utilização dos dados agregados para caracterização de pessoas singulares (doravante designada de "Fase 2")

Figura 5 - Utilização dos dados agregados para caracterização de pessoas singulares (Fase 2)



Fonte: Elaboração própria

A Figura 4 – Recolha e Agregação de Dados representa a Fase 1 do projeto onde os dados pessoais são recolhidos de vários titulares dos dados nos diversos serviços disponibilizados pelos municípios. Nesta fase, os dados são recolhidos, por regra, diretamente do titular dos dados e permitem a identificação do mesmo.

No sentido de proteger a privacidade dos titulares dos dados é utilizada a técnica de agregação de dados onde são obtidas informações sumarizadas. Estas informações sumarizadas não contemplam dados que identifiquem pessoas singulares, constituindo dados de um determinado grupo.

Como exemplo ilustrativo desta primeira fase temos a recolha de dados de consumos de água de uma determinada localidade por habitante para fins de prestação de serviços de fornecimento de água. À partida, estes dados seriam tratados única e exclusivamente para a referida finalidade (i.e., execução do contrato de prestação de serviços de fornecimento de água), contudo, com o

objetivo de otimizar a presente recolha, os dados vão ser objeto de agregação. Os dados obtidos através da técnica de agregação de dados constituem uma referência abstrata que caracteriza todo o conjunto de habitantes da referida localidade.

A Figura 5 - Utilização dos dados agregados para caracterização de pessoas singulares (Fase 2) - representa a fase do projeto onde os dados agregados (i.e., que constituem uma referência abstrata) são utilizados para efeitos de caracterização de uma pessoa singular. Com as referências abstratas obtidas, que caracterizam um grupo de pessoas, pretende-se inferir potenciais comportamentos / decisões dos titulares dos dados. Ora, tendo em consideração que são obtidas referências abstratas, quando relacionadas com pessoas singulares, em regra representam meras probabilidades.

Pegando no exemplo acima apresentado, a referência abstrata obtida, que caracteriza os consumos de água de todo o conjunto de habitantes da referida localidade, vai ser relacionada com pessoas singulares no sentido de procurar prever o seu consumo de água. Através desta relação entre a referência abstrata obtida e uma pessoa singular, em regra, vamos obter uma probabilidade de uma pessoa A ter um determinado consumo.

6.2. Subsunção dos dados agregados ao conceito de dados pessoais

Cumpre agora analisar se os dados agregados (i.e., dados sumarizados) são ou não subsumíveis no conceito de dados pessoais constante do n.º 1 do art. 4.º do Regulamento Geral de Proteção de Dados. Para tal, será necessária a análise dos 5 elementos constantes da definição de dados pessoais, em particular: i) informação; ii) relativa a; iii) uma pessoa singular; iv) identificada; e v) identificável.

Como vimos anteriormente o Regulamento Geral de Proteção de Dados procura oferecer um conceito amplo de dados pessoais, começando por referir que dados pessoais são informações. O conceito de informação não se encontra detalhado no RGPD, sendo comummente entendido como um conjunto de dados organizados que proporcionam sentido e valor para o recetor. Tratando-se de dados agregados constituem necessariamente dados organizados e tratados

que têm valor e sentido quando interpretados, pelo que concluímos que estes dados são considerados informações.

Concluído que os dados agregados são informações cumpre analisar se estas informações são relativas a uma pessoa singular[45]. Para que as informações sejam consideradas dados pessoais é necessário que exista uma relação entre a informação e um determinado sujeito.

Como é referido no n.º 1 do art.1.º, do Regulamento Geral de Proteção de Dados[46], as regras estabelecidas neste visam a proteção de pessoas singulares. Um dos principais objetivos do Regulamento Geral de Proteção de Dados prende-se com a devolução do controlo às pessoas singulares dos seus próprios dados pessoais e, o coração do Regulamento, são os direitos, liberdades e garantias da pessoa singular.

No que concerne aos dados obtidos através da técnica de agregação é necessário compreender a quem pertencem os atributos objeto de agregação, isto é:

> a) A técnica da agregação pode ser aplicada a vários atributos de uma pessoa singular. Nesta situação, os dados agregados obtidos e quando relacionados com a pessoa singular em causa, configuram dados pessoais na medida em que são "*relativos a uma pessoa singular*" sendo esta pessoa singular "*identificada*". Neste caso, quando os atributos objeto de agregação respeitam a uma pessoa singular, os dados agregados constituem dados pessoais, nos termos e para os efeitos do n.º 1 do art. 4.º do RGPD.
>
> b) Quando a técnica da agregação é aplicada aos atributos partilhados por x utilizadores, os dados resultantes da agregação passam a ser dados referentes a um grupo. Neste caso, deixa de ser possível selecionar uma pessoa dentro de um grupo de x pessoas[47]. Ora, se os dados deixam de ser dados relativos a pessoas

45 Cfr. Considerando 14 conjugado com o art. 3.º do RGPD.

46 "O presente regulamento estabelece as regras relativas à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados.", art. 1.º, n.º 1 do Regulamento Geral de Proteção de Dados.

47 GT 29, Parecer 05/2014, cit., 18.

singulares e passam a ser dados relativos a grupos de pessoas, não são subsumíveis ao conceito de dados pessoais constante do RGPD.

Nesta última situação, poderá colocar-se a questão de o número de pessoas (x) ser um valor baixo. É importante notar que, quanto maior é o número de pessoas, maior é a privacidade e menor é a taxa de sucesso em caso de tentativa de re-identificação. No entanto, tendo em consideração que o objeto de análise é a possibilidade de subsunção ao conceito de dados pessoais que nos é apresentado pelo RGPD, desde que o x seja superior ou igual a 2, os dados são considerados dados anónimos.

Adicionalmente, é necessário analisar se uma probabilidade de determinado dado agregado pertencer a uma determinada pessoa é ou não é um dado pessoal. Atualmente, já existem variados estudos que permitem o cálculo da probabilidade de determinado *data set* agregado ser objeto de re-identificação. A questão que se coloca neste âmbito é se obtivermos uma probabilidade de 10% ou de 90% de determinado dado agregado pertencer a um perfil de uma pessoa singular e identificá-la, é ou não dado pessoal.

Para verificarmos a subsunção ao conceito será necessário relembrar que o Regulamento considera dados pessoais as informações relativas a uma pessoa singular identificada ou identificável, sendo identificável uma pessoa singular que possa ser identificada, direta ou indiretamente. A letra da lei é clara quando refere que, para que determina informação seja considerada um dado pessoal, é necessário que a mesma permita identificar uma pessoa direta ou indiretamente.

Assim sendo, nos casos em que exista uma determinada probabilidade de um perfil pertencer a uma determinada pessoa, não poderá ser considerado um dado pessoal, quer se trate de uma probabilidade de 10% ou de 90%. Isto porque, independentemente da probabilidade em causa, é sempre um dado relativo a um grupo de pessoas. Para além de ser um dado de um grupo de pessoas, não constituí uma informação relativa a uma pessoa "*identificada ou identificável*".

Neste sentido, concluímos que sempre que tenhamos uma probabilidade de identificação, independentemente de qual for, nunca será subsumível ao

conceito de dados pessoais constante do Regulamento Geral de Proteção de Dados.

Ora, a partir do momento em que deixamos de ter uma probabilidade e passamos a ter uma certeza na identificação de uma pessoa singular, os dados agregados passam a ser dados pessoais para efeitos do Regulamento Geral de Proteção de Dados.

Assim sendo, os dados agregados são considerados dados anónimos quando calculados com base em atributos de um grupo de pessoas e quando existam apenas probabilidades de identificação de uma pessoa singular.

6.3. Os tratamentos de dados no âmbito das Smart Cities

Passemos agora para a análise dos tratamentos de dados no âmbito das Smart Cities, ou seja, no âmbito das fases acima identificadas (i.e., Fase 1 e Fase 2). O objetivo pretendido é compreender em que momentos estamos perante dados pessoais, nos termos do n.º 1 do art. 4.º do Regulamento Geral de Proteção de Dados.

A primeira fase do projeto é caracterizada pela recolha de dados pessoais dos cidadãos que é levada a cabo pelos municípios, em particular, no âmbito da gestão dos serviços públicos. Relativamente a este primeiro momento da primeira fase do projeto, não existem dúvidas de que estamos perante dados subsumíveis ao conceito de dados pessoais. As informações recolhidas por parte dos municípios respeitam a pessoas singulares, sendo possível identificar diretamente as mesmas.

De seguida, ainda dentro desta primeira fase, os dados respeitantes a pessoas singulares são objeto de agregação. Através desta técnica, os dados são agregados por x indivíduos, constituindo sempre dados relativos a um grupo de pessoas. Ora, se são dados relativos a um grupo de pessoas não respeitam uma pessoa singular.

No contexto das *Smart Cities*, os dados agregados respeitam sempre a grupos de pessoas, não se colocando a questão levantada de se tratarem de vários atributos pertencentes a uma pessoa.

Na segunda fase, ilustrada na figura 5, os dados obtidos através do processo de agregação na fase 1, são relacionados com pessoas singulares. Exemplo desta situação ocorre quando se obtém uma referência abstrata relativa ao consumo de água de um determinado grupo de pessoas e esse consumo é relacionado com uma determinada pessoa singular de forma a obter uma probabilidade da mesma ter o referido consumo.

Neste caso é necessário distinguir as situações em que, através da relação entre os dados agregados e uma pessoa singular, se obtém uma probabilidade ou uma certeza sobre a mesma. Como vimos anteriormente, sempre que se obtenham probabilidades sobre um dado poder pertencer a determinada pessoa, independentemente de qual for a probabilidade, nunca são dados pessoais. Estes dados não são pessoais na medida em que não são "*dados relativos a uma pessoa singular*". Contudo, nas situações em que obtemos certezas sobre uma pessoa singular, passamos a ter "*dados relativos a uma pessoa singular identificada*".

Em suma, analisadas as duas fases do projeto, apenas no momento inicial de recolha de dados estaremos perante dados pessoais. O único momento em que voltamos a ter dados pessoais é quando relacionamos os dados agregados com as pessoas singulares e obtermos certezas sobre as mesmas.

7. Conclusão

O presente estudo foi despoletado pela verificação atual de uma recolha massiva de dados pessoais dos munícipes, bem como o desenvolvimento de tecnologias de informação que permitem ainda uma recolha de maior qualidade, quantidade e variedade. Através desta verificação, foi percecionada a necessidade de uma solução que visasse otimizar a recolha de dados, de forma a prestar um melhor serviço público.

Neste contexto, através do presente estudo, procurámos encontrar um critério legal que permitisse concluir a partir de que momento os dados agregados são considerados dados anónimos[48]. A procura desta solução visa

48 Cfr. art. 1.º, 2.º e 3.º e considerando 26 do Regulamento Geral de Proteção de Dados.

permitir que as cidades utilizem os dados recolhidos sem colocar em causa a privacidade dos cidadãos.

Nesta análise concluímos que os dados agregados são dados anónimos quando sejam calculados com base em atributos de um grupo de pessoas e quando exista apena uma probabilidade de identificação de uma pessoa singular.

Concluímos ainda que, através deste critério será possível que as cidades inteligentes otimizem a utilidade dos dados já recolhidos, não sendo necessário o cumprimento do Regulamento Geral de Proteção de Dados e não ficando sujeito ao princípio da limitação das finalidades.