Abstract: This paper explores the relation between the European rules on data protection and the commercialization of data. After contextualising the ongoing data revolution, it is argued that the practice of commercializing data may raise problematic issues under data protection law in two fundamental respects. First, it is often unclear whether data protection law should apply to begin with (‘the “if” problem’). Secondly, when and if data protection law applies, it is frequently hard to determine how it should apply (‘the “how” problem’). As we will observe, both the irreducibly dynamic nature of data and the conceptual blurriness of our legal discourse risk setting data protection law and trade in data on a baleful collision course.

Résumé: Cet article explore la relation entre les règles européennes sur la protection des données et la commercialisation des données. Après avoir contextualisé la révolution des données en cours, il est avancé que la pratique de commercialisation des données peut soulever des problèmes au regard de la législation sur la protection des données à deux égards fondamentaux. Premièrement, il est souvent difficile de savoir si la loi sur la protection des données doit s’appliquer au départ («le problème du « si »). Deuxièmement, quand et si la loi sur la protection des données s’applique, il est souvent difficile de déterminer comment elle doit s’appliquer («le problème du « comment »). Comme nous le verrons, tant la nature irréductiblement dynamique des données que le flou conceptuel de notre discours juridique risquent de définir le droit de la protection des données et le commerce des données sur une trajectoire de collision funeste.

Zusammenfassung: In diesem Artikel wird der Zusammenhang zwischen den europäischen Datenschutzbestimmungen und der Kommerzialisierung von Daten untersucht. Nach der Kontextualisierung der laufenden Datenrevolution wird ar-
gumentiert, dass die Praxis der Kommerzialisierung von Daten in zweierlei Hin-
sicht problematische Aspekte im Rahmen des Datenschutzrechts aufwerfen kann. 
Erstens ist zunächst oft unklar, ob das Datenschutzrecht anwendbar ist („das Ob-
Problem“). Zweitens, sofern das Datenschutzrecht gilt, ist es häufig schwierig zu 
bestimmen, wie die Datenschutzgesetze anzuwenden sind („das Wie-Problem“). 
Wie wir feststellen werden, riskieren sowohl die nicht reduzierbare dynamische 
Natur von Daten als auch die konzeptionelle Unschärfe unseres Rechtsdiskurses, 
das Datenschutzgesetz und den Handel mit Daten in Richtung eines unheilvollen 
Kollisionskurs zu lenken.

1 Introduction

The World is in the middle of a data revolution: it is no overstatement to say so. 
Data are driving key shifts in our cultural, socio-economic, and political practices, 
creeping into classrooms and boardrooms alike.¹ In B2C transactions, digital con-
tent and digital services are often paid for with personal data, with the latter being 
increasingly seen as a ‘tradable assets’ and accepted as counter-performance.² We 
cannot practically escape this revolution nor gloss it over in our theorising: the 
only promising way forward is to try and understand it. Some have been exuber-
ant enough to hail data as the new ‘oil’ of the economy;³ many have toiled away at 
graphs and statistical models to try and predict the future growth of our data en-
vIRONMENT;⁴ and most agree that data are an increasingly valuable asset in today’s 
world.⁵

² S. Lohsse, R. Schulze and D. Staudenmayer, ‘Data as Counter-Performance – Contract Law 2.0? An Introduction’, in S. Lohsse, R. Schulze and D. Staudenmayer (eds), Data as Counter-Performance – Contract Law 2.0? Münster Colloquia on EU Law and the Digital Economy V (Baden-Baden: Nomos, 2020) 10. For an insightful overview of the ‘data as counter-performance’ debate, see Lohsse, Schulze and Staudenmayer (eds), this n above.
³ ‘The world’s most valuable resource is no longer oil, but data’ The Economist (London, 7 May 2017).
⁴ Commission, ‘Towards thriving data-driven economy’ COM(2014) 442 final (predicting that the data economy will be worth € 739bn or 4 % of the EU GDP by 2020); D. Tang et al, ‘Value of data: the dawn the of data marketplace’ (Accenture, 2018) <https://www.accenture.com/us-en/insights/high-tech/dawn-of-data-marketplace> (suggesting that data marketplaces may be worth more than € 3.6tn by 2030).
⁵ Many M&A transactions have been carried out partly or wholly for data-related purposes: eg, data played a substantial role in the Microsoft-LinkedIn, Facebook-WhatsApp and Google-Double-
Similarly, however, no one denies that data are a Janus-faced creature and that a cesspool of dangers lurks behind their tempting Eden of opportunities. Privacy harms, algorithmic discrimination, profiling, behavioural targeting and cyberattacks: these are but a few of the issues we confront anytime we thoughtlessly tick the ‘I agree’ box under a website’s privacy notice and agree to the processing of our data. Before delving deeper into commercialization of data, let us step back and revise the fundamentals.

2 What are Data?

In both legislation and scholarly literature, data are often defined by reference to the notion of ‘information’. Unsurprisingly, that of ‘information’ is a multifaceted, if not inherently contested, concept; and its relation to data more problematic than it might seem at first. More specifically, some use the term ‘data’ to refer to both syntactic and semantic information. At the syntactic level, information is a mere aggregate of physical signs and symbols: alphabetic letters, soundwaves emitted by our vocal cords as they vibrate, strings of binary code, and the likes. On the other hand, semantic information is the meaning which either human agents or machines can derive from interpreting signs. We might think of encrypted messages to see how these two levels of information can crucially dif-

---


8 See art 4(1) GDPR (‘personal data are information’).


fer. Undoubtedly, an encrypted message is syntactic information.\textsuperscript{12} It may be either a 256-bit string of 0s and 1s if the message is encrypted via a SHA-256 hash function; or an unusually-arranged sequence of alphabetic signs if a plaintext is encrypted via Caesar cipher cryptography.\textsuperscript{13} However, we can extrapolate no meaning out of an encrypted message unless we decrypt it first.\textsuperscript{14} As such, an encrypted message is syntactic information, but has not semantic content: stated otherwise, it has no intelligible meaning and carries nothing over and above a physical cluster of signs and symbols. As we shall see in the paragraphs below, both statutory instruments and the scholarly literature frequently refer to data just as ‘information’.\textsuperscript{15} However, it is crucial to bear the semantic-syntactic divide in mind. As we will see below, the term ‘data’ may often be used to refer to both signs (at the syntactic level) and their meaning (at the semantic level). But at times, a distinction can be drawn between ‘data’ as mere signs (syntactic element) and ‘information’ as meaning (semantic element). To avoid puzzlement and ambiguity, we had better be aware of the different senses in which ‘data’ is being used each time.\textsuperscript{16}

When seen as mere units of information, data are nothing new under the sun; nor could they ever be. As hinted above, we and the environment around us are incessantly churning data out. Anytime a leaf falls off a branch or our stomach growls, a signal is being transmitted across the physical space and new data are being created.\textsuperscript{17} However, something else has changed lately to reconfigure our relation to data. If this had not been the case, our whole talk of a ‘data economy’\textsuperscript{18}

\textsuperscript{12} For a similar stance on encrypted messages, see M. Boisot and A. Canals, ‘Data, information and knowledge: have we got it right?’ (2004) \textit{Journal of Evolutionary Economics} 43.
\textsuperscript{13} A hash function is a function that maps input of any arbitrary size into a fixed-size output. In the case of a SHA-256 hash function, the input can be any data from Dante’s \textit{Comedia} to a single digit, and the output will always be a string of 256 bits. On the other hand, the Caesar cipher is a widely-known encryption technique that substitutes each letter of a plaintext with another letter. Usually, the substituting letter is a fixed number of positions down the alphabet.
\textsuperscript{14} It may be argued that an encrypted message has a ‘hidden’ meaning. However, we can only see it after decrypting it. However, the decrypted text is different from the encrypted one: we only extract meaning from the decrypted message, not from the encrypted one.
\textsuperscript{15} GDPR, art 4(1); GDPR, recital 26; see also G. Malgieri, ‘Property and (Intellectual) Ownership of Consumers’ Information: A New Taxonomy for Personal Data’ (2016) \textit{Privacy in Germany} 133.
\textsuperscript{16} For a view keeping information and data as two distinct notions – i.e., data as only signs and symbols; and information as the meaning or significance that can be derived from them – see Floridi, n 10 above.
\textsuperscript{17} See Floridi, n 10 above.
and of a ‘data-driven society’\textsuperscript{19} would be intolerably nonsensical. Let us try and unpack the content of this shift.

3 Get it, Keep it, Use it: the Changes in the Data World

Put simply, what has radically improved over the last few years is our capacity to collect, store and analyse data. We shall look at each of these in turn. Firstly, the digital era ushered in an unprecedented ability to harvest data.\textsuperscript{20} Scrolling down a website, inputting queries into a search engine, making online payments and interacting with one’s peers on social media platforms: these and almost every other online operation of ours leaves a digital trail behind and generates data points ready for collection.\textsuperscript{21} Furthermore, the rise of the Internet of Things has spectacularly increased our data-collection capabilities.\textsuperscript{22} In a nutshell, the Internet of Things (IoT) can be defined as an ecosystem of interconnected digital devices.\textsuperscript{23} Its rise has not only challenged existing product liability and safety regimes, but also reshaped existing data-collection practices.\textsuperscript{24} In particular, we no longer have to sit in front of a desktop and drag our mouse around to generate data. By simply wearing a smart watch or making use of a connected fridge, we can leave a

\begin{itemize}
\item [24] For the challenges that the Internet of Things and Artificial Intelligence pose to established product liability and product safety frameworks, see S. Lohsse, R. Schulze and D. Staudenmayer, Liability for Artificial Intelligence and the Internet of Things: Münster Colloquia on EU Law and the Digital Economy IV (Baden-Baden: Nomos, 2019).
\end{itemize}
digital footprint behind and generate a disconcerting number of data for organisations to pick up.25

Yet being able to collect mammoth amounts of data would be almost useless if we could not store them and retrieve them efficiently. Lo and behold, our storage and retrieval capabilities have also greatly improved over the last few years. In particular, two innovations are here worth mentioning: cloud storage and data virtualization. A cloud storage is a cloud-computing system that involves storing data on a remote server. A third party managing and operating the storage infrastructure delivers the data-storing service over the Internet on demand and often through a pay-as-you-go model. This eliminates the need for many to arrange and operate burdensome physical storage infrastructures. As a natural counterpart to efficient storage, data virtualization has enabled users to retrieve and access data without having to care where the data are stored, in what form they are coded or how their servers are run. For the user, accessing data from multiple sources and in multiple formats will feel like accessing a single unified large database.26

Nevertheless, the most tremendous shift in our data environment has likely come from an unexampled ability to process data and use them to inform our predictions. Admittedly, predictive analytics has existed since the dawn of time. Past events have always informed our prognostics and educated our guesses.27 Today, however, the increasingly-larger size of our datasets, the advance in our technology and computing power and the development of statistical analysis and data mining, among others, have made it far easier to infer patterns from our data and come up with well-informed and reliable predictions.28

4 A New Opportunity for Businesses: Commercialising Data

Against this background, it is no surprise that data are incredibly useful for organizations and that every business has been forced to turn into a ‘data business’

---

27 Helveston, n 1 above.
somehow. If we were to try and categorise the unlimited ways businesses can valorise their data, we would ultimately come up with two types of use: internal and external uses. Internally, businesses can analyse data to discern patterns in customers’ or clients’ behaviours, inform their strategic decisions and expand their reach, to name but a few examples. Externally, firms can decide to trade their data, in either their raw or processed form. For instance, a genomics company collecting health data from vials of saliva may wish to monetise them by giving a pharmaceutical company access to their dataset. This external use of data is what we shall refer to as ‘commercialisation of data’ (or ‘data trade’): namely, the set of business-to-business data transactions (‘B2B transactions’) and the resulting extraction of monetary value from them. Surely, a data subject paying for the provision of digital content or services with her data is also ‘trading’ or ‘commercialising’ data in the relevant sense. However, consumer-to-business data transactions (‘C2B transactions’) per se fall beyond the scope of our paper.

A business intending to commercialize consumers’ data and trade them with a third party will encounter a panoply of legal issues. Yet the very first concern for any such business (which we shall refer to as ‘data-trader’) will be to meet their obligations under data protection law. As we shall go on to argue, this apparently unproblematic question raises thorny and often unsurmountable issues in at least


30 Ibid 29.


32 Du Boulay et al, n 29 above, 29.

33 L. Matsakis, ‘The WIRED Guide to Your Personal Data (and Who Is Using It)’ (Wired, 15 February 2019) <https://www.wired.com/story/wired-guide-personal-data-collection/>. 34 ‘Monetisation’ is often used interchangeably with ‘commercialisation’. However, monetisation is wider in scope: data are monetised anytime some monetary value is extracted from them. Therefore, even a mere analysis is a form of monetisation. 35 ‘Commercialisation’ is a wider notion than mere business-to-business data trade (the focus of our paper): a data subject trading her data for digital content or services is commercialising (and trading) personal data too. If digital content or services are being paid for with personal data, the trader will be caught under the Digital Content Directive: see Directive (EU) 2019/770 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning contracts for the supply of digital content and digital services [2019] OJEC L136, art 3. For an analysis of the scope of the DCD, see D. Staudenmayer, ‘Art 3’, in R. Schulze and D. Staudenmayer, EU Digital Law: Article-by-Article Commentary (Bloomsbury Academic, 2020) 57 et seq. For an exploration of the notion of ‘data as consideration’, see also C. Langhanke and M. Schmidt-Kessel, ‘Consumer Data as Consideration’ (2015) 4 Journal of European Consumer and Market Law 219 (pre-dating the Digital Content Directive).
two fundamental respects. Firstly, it is often difficult to determine whether data protection law applies to begin with, given the obscurity bedevilling the notion of ‘personal data’. Secondly, when and if data protection law applies, it is hard to figure out how it should apply. We shall examine each of these issues in turn from a European law perspective.

5 The ‘If’ Problem

5.1 Personal Data: an ‘Essentially Contested’ Concept

The data protection regime in both the European Union (EU) and the European Economic Area (EEA) is enshrined in the now-infamous General Data Protection Regulation (hereinafter, the ‘GDPR’), which was adopted on the 14 April 2016 and came into force on the 25 May 2018. The GDPR lays out a legal framework for processing personal data within the EU and the EEA. Under the GDPR, any ‘data controller’ (someone determining the purposes of the processing) must process data ‘fairly, lawfully and in a transparent manner’ and must observe the principle of accountability together with that of purpose limitation, data minimisation, accuracy, storage limitation, integrity and confidentiality.

Crucially, the GDPR only regulates the processing of personal data. That is to say, non-personal data, otherwise known as ‘anonymous’ data, fall outside its remit. For our purposes, it is therefore imperative to unpack the definition of ‘personal data’ under the GDPR and clarify its scope. As our intuition would suggest, commercialising personal data will be ostensibly more burdensome than

---

38 General Data Protection Regulation, art 4(7).
39 General Data Protection Regulation, art 5.
merely trading anonymous information. Unfortunately, despite its cardinal role within the GDPR’s legal architecture, the concept of personal data remains astonishingly elusive.\textsuperscript{41} It is to this very concept that we now turn.

**5.2 ‘Omnis Definitio Periculosa Est’\textsuperscript{42}\textsuperscript{43}\textsuperscript{44}\textsuperscript{45}**

Article 4(1) GDPR defines personal data as:

\begin{quote}
‘any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.’\textsuperscript{43}
\end{quote}

Reasoning \textit{a contrario}, we can immediately discern two types of information likely to escape this definition: (1) data relating to legal persons; and (2) data relating to a deceased.\textsuperscript{44} Similarly, data informing us that ‘the weather is rainy in London’ or that ‘water boils at 100°C’ seem \textit{prima facie} outside the scope of Article 4(1).\textsuperscript{45} On a closer inspection, however, the definition provided by Article 4(1) is far more slippery than it may seem.

**5.3 ‘Information’**

Firstly, data are defined as ‘information’.\textsuperscript{46} As seen above, the notion of ‘information’ may have a two-fold interpretation: (a) it may encompass both the syntactic level (eg, bits and bytes) \textit{and} the semantic level (eg, the meaning or ‘message’ a machine can extract from binary code); or (b) it may be restricted to the semantic level (so that ‘information’ would imply ‘meaning’, and the concept of ‘meaning-
less information’ would be a contradiction).\textsuperscript{47} Some suggest that data must carry some ‘informational value’ to satisfy the definition under Article 4(1), perhaps implying that they should carry some meaning.\textsuperscript{48} Again, this suggestion relies on a distinction often drawn in information theory between ‘data’ seen as mere symbols (syntactic level), and ‘information’ restrictively seen as the \textit{meaning} of such symbols (semantic level).\textsuperscript{49} On this view, equating data with information is tantamount to saying that data should mean something. Yet this distinction is highly problematic: encrypted messages only have syntax and carry no meaning \textit{per se}. Hence if we take ‘meaning’ to be synonymous with ‘informational value’, encrypted messages would have no ‘informational value’ and should run afoul of Article 4(1). To see the problem more vividly, consider a scenario taken from outside the realm of the GDPR. During a class, Alice hands over to her classmate Charles a piece of paper with a line of writing reading: ‘Ere lv loo’. The syntactic level in this line of text is just the sequence of letters as written on the paper (‘E’ followed by ‘r’, followed by ‘e’, etc); the semantic level is the \textit{meaning} of such sequence. Yet there is no intelligible meaning in the plaintext: the line \textit{in itself} is nonsensical. However, if we decode it with a Caesar-cipher cryptography (shifting each letter three places up the alphabet), we see that a new message crystallises: ‘Bob is ill’. \textit{This} line of text has a new syntactic configuration (a different sequence of letters) and a plain and intelligible meaning. If we require data to have meaning, then ‘Ere lv loo’ \textit{in itself} would not count as data as it fails to mean anything: in general, no encrypted message would. This conclusion would fly in the face of the GDPR and should be firmly resisted.\textsuperscript{50} If we swap Alice for a corporation circulating encrypted material on their employees’ financial history, we would not want them to rely on a niche sophistry of information theory to escape the GDPR and argue that any such encrypted material is meaningless in itself and is therefore no ‘information’. Still, the relation between data and information remains a vexed one to flesh out.\textsuperscript{51}

\textsuperscript{47} For the concept of ‘semantic information’, see Floridi, n 10 above.
\textsuperscript{48} See M. Finck and F. Pallas, ‘They who must not be identified—distinguishing personal from non-personal data under the GDPR’ (2020) \textit{International Data Privacy Law} 3 (noting ‘that Article 4 (1) GDPR refers to ‘information’ rather than just data, indicating that the data appears to require some informational value’, 13).
\textsuperscript{50} See Boisot and Canals, n 12 above, 43 (noting that in encrypted messages ‘while the data itself can be made “public” and hence freely available, only those in possession of the “key” are in a position to extract information from it (Singh, 1999). Cryptography, in effect, exploits the deep differences between data and information’).
\textsuperscript{51} Janeček, n 22 above, 1042.
Interpreting the similarly-worded definition of ‘personal data’ under the former Data Protection Directive (or ‘DPD’), the Article 29 Working Party (now European Data Protection Board) opined that information can be both objective and subjective. This was confirmed by the CJEU in Nowak. The salient issue before the court in that case was whether an examination script and an examiner’s feedback on it could constitute ‘personal data’ despite their subjectivity. Confirming the conceptual breadth of the notion of ‘personal data’ (albeit under the former DPD), the CJEU held that:

‘the expression ‘any information’ [...] reflects the aim of the EU legislature to assign a wide scope to [the notion of personal data], which is not restricted to information that is sensitive or private, but potentially encompasses all kinds of information, not only objective but also subjective, in the form of opinions and assessments [emphasis added].’

Therefore, the notion of ‘personal data’ under the GDPR will similarly embrace not only objective facts but also any subjective statement (be it an assessment, an opinion or possibly even a mere speculation) provided that it ‘relates to’ an identified or identifiable natural person. It is to the meaning of ‘relating to’ that we now turn.

5.4 ‘Relating to’

Secondly, information must ‘relate to’ an identified or identifiable natural person. ‘Relating to’ may refer to three things: information that is about a natural person (a ‘content’ test); information that is meant to affect a natural person (a ‘purpose’ test); information which may result in a natural person being affected (an ‘impact’ test). Endorsed by the A29 Working Party, this threefold interpretation of the phrase ‘relating to’ proves highly problematic. In principle, any information, including mere weather readings, may be intended to affect a natural person (satisfying the ‘purpose’ test); or may result in a natural person being impacted (meet-
ing the ‘impact’ test). In an eerily-paradoxical example provided by Nadezhda Purtova, weather readings collected in a smart-city living lab to influence the behaviours of its inhabitants are data intended, or likely, to affect people: 58 by ‘relating to’ the city’s residents in both their purpose and impact, weather readings should then count as ‘personal data’ too.59 This is why Purtova goes on to vibrantly warn that data protection law could risk becoming ‘the law of everything’.60

Note, however, that others are more optimistic. In particular, some scholars emphasise that ‘personal data’ must still relate to ‘an identifiable natural person’ and not just to ‘people’ as an undefined category.61 By way of example, weather-pattern data, if used to affect the functioning of a smart city, would not relate to any specific individual and should not count as ‘personal data’ within the meaning of Article 4(1) GDPR; but if used by an insurance company to calculate premiums for a specific insured (eg, based on the precipitation rates and the resulting likelihood of rainfall-related accidents in her local area), weather readings would relate to an identifiable natural person and should therefore qualify as personal data.62 In other words, the need for the natural person – to whom the data relate – to be either ‘identified’ or ‘identifiable’ could serve as a curbing mechanism, neutralizing the breadth of the ‘relating to’ element and preventing data protection law from becoming the law of everything.

5.5 ‘Identified or Identifiable Natural Person’

The notion of identifiability plays a cardinal role in our understanding of ‘personal data’, both in European data protection law and in almost any other data protection regime.63 In the GDPR, personal data are defined as data relating to ‘an identified or identifiable natural person [emphasis added]’.64 Recital 26 provides further guidance on the notion of identifiability. It alerts us that:

58 Purtova, n 45 above, 40.
60 Ibid 40.
64 General Data Protection Regulation, art 4(1).
‘[...] To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments [...]’

Albeit not legally binding, recitals are an authoritative source of insight and should inform our interpretation of the statutory text. There are at least three items that we may wish to pick up on in Recital 26: a (i) risk element; (ii) a third-party element; and an (iii) ongoing-duty element.

5.6 The Risk Element: a Risk-Based Approach to Identifiability

Glancing over the text of Recital 26, we may infer that the GDPR invites a risk-based approach to anonymisation. In other words, data are considered anonymous if it is unlikely that they can be re-identified: the ‘reasonably likely means’ test seemingly bears out this risk-based interpretation of ‘identifiability’. Under this approach, if a data controller D holds a dataset and it is reasonably unlikely that D could re-identify data subjects within her dataset, D’s data may well count as anonymous. However, there is more to say than meets the eye. When interpreting the notion of identifiability as found in the Data Protection Directive, the A29 Working Party seemed to endorse a far more stringent no-risk approach to anonymisation. In its ‘Opinion on Anonymisation Techniques’, it suggested that the de-identification process must be ‘irreversible’ before data can be deemed anonymous and thereby exempted from data protection legislation. In an unfortunate oxymoron, the Opinion referred to the need for a ‘reasonable impossibility’ of re-identification in order to conclude that data have become non-personal. To go back to the example above: under a zero-risk approach, it must not only be reasonably unlikely that data controller D could re-identify the data subjects within her dataset; if her data are to count as anonymous, it must be impossible for her to do so.

65 General Data Protection Regulation, recital 26.
67 Art 29 Working Party, Opinion 05/2014 on Anonymisation Techniques (WP 216) 0829/14/EN, 7.
68 Ibid 8.
However, a zero-risk approach to identifiability should be dismissed. Not only does a textual analysis of Recital 26 GDPR fly in the face of a zero-risk approach to anonymisation; demanding that the risk of identifiability be null would also be ill-advised as a matter of policy. Over the years, it has become increasingly clear that the ideal of an ‘irreversible anonymisation’ is a chimera, and that no de-identification can ever eliminate the risk of re-identification.69 History is sadly littered with examples of reversed anonymisation. The AOL data release, the Netflix Prize Data Study, and the Massachusetts health database re-identification are an infamous triad of failed attempts at anonymisation, demonstrating how many instances of assumedly-effective de-identification could eventually be reversed.70 Nowadays, the ‘broken promise of privacy’ is acknowledged and sorrowed over by legal scholars and computer scientists alike.71 Thanks to advances in ‘reidentification science’, we have learnt that ‘data can be either useful or perfectly anonymous but never both’.72 Furthermore, that the GDPR approach to identifiability should be more nuanced and relative than the A29WP’s Opinion recommends is also indirectly confirmed by the CJEU’s seminal ruling in Breyer.73 It is to this judgement that we now turn.

5.7 The Third-Party Element: a Relative Approach to Identifiability

The test of identifiability under Recital 26 GDPR stretches beyond the data controller. When considering the means ‘reasonably likely to be used to identify’ a data subject, it considers not just the wherewithal at the data controller’s disposal, but also any means reasonably likely to be used ‘by another person’.74 Consider a data controller D holding a partial dataset, eg, one with dynamic IP addresses (where dynamic IP addresses change each time there is a new Internet connection). D could only identify its data-subjects by combining its dataset with that held by a third party: eg, Internet Service Provider in D’s case. We will refer to

70 Ibid 1717–1722.
72 Ohm, n 69 above, 1704.
73 C-582/14 Patrick Breyer v Bundesrepublik Deutschland [2016] EU:C:2016:779 (‘Breyer’).
74 General Data Protection Regulation, recital 26.
this third-party-held material enabling re-identification as ‘additional re-identifying information’ or ‘ARI’.75 A crucial issue in these scenarios is whether the mere fact that the ARI exists somewhere in the world is sufficient to turn the non-identified data in the controller’s partial dataset into personal data. The wording in Recital 26 is silent, if not perplexing, in this respect: it merely refers to the means reasonably likely to be used ‘by another person’, without specifying if we should take that to mean ‘any person’ in the world (what would effectively amount to an ‘absolute approach’ to identifiability).76 If Recital 26 was given such a broad interpretation, there would inevitably be an overkill. In principle, if we had to consider any potential ARI in the hands of hypothetically anyone across the globe, we would, and indeed should, conclude that any de-identified information, including highly-aggregated one, could be re-identified. However, the CJEU in Breyer seemed to dispel this dramatically-broad interpretation of Recital 26.

In Breyer, the European Court of Justice had to consider whether dynamic IP addresses stored on a website provider were personal data. It held that they were in this case.77 Although the data controller could not identify anyone based on the data at their disposal, the Internet Service Provider (ISP) had additional information which it could use to track and identify users associated with each dynamic IP address.78 Admittedly, if taken at face value, the judgement may be seen to give Recital 26 an expansive interpretation; and indeed, many have interpreted the ruling as entrenching that very ‘absolute’ approach to identifiability we were anxious to stave off a few lines above.79 On a closer inspection, however, the approach to identifiability endorsed by the CJEU in Breyer is far more nuanced and ‘relative’ than it might appear.

Firstly, the CJEU reiterated the need for a contextual appraisal of all the relevant circumstances of a case when assessing the likelihood of re-identification; and secondly, it held that the legality of any purported avenue of re-identification should also be considered as a relevant factor in assessing identifiability.80 If

---

75 The examples provided are inspired by the facts in Breyer.
77 Breyer, n 73 above, [47].
78 For a useful explanation of dynamic IP address, see Purtova, n 45 above, (‘a dynamic IP address changes each time there is a new Internet connection, and does not allow a link to be established, ‘through files accessible to the public’, between a given computer and the Internet connection’, at 65).
79 Purtova, n 45 above, 62. Although the judgement refers to recital 23 of the former Data Protection Directive, this has a markedly-similar wording to recital 26 GDPR: Breyer can be considered authority for both.
80 Breyer, n 73 above, [45]–[49].
crossing two datasets for re-identification purposes is an act prohibited under national law, that should be ruled out as a means not ‘reasonably likely to be used to identify’ a data subject.\textsuperscript{81} This ‘lawfulness’ factor was crucial on the facts of Breyer itself. In that case, the Internet Service Provider was not allowed under German law to release the ARI in ordinary circumstances. Nevertheless, the website provider (holding the partial dataset of dynamic IP addresses) could have reached out to competent authorities in the event of a cyberattack; and the authorities, in turn, would have been legally entitled to request the ARI from the Internet Service Provider in order to commence criminal proceedings against any identified attacker. In other words, there was no blanket ban on securing the ARI, nor any legal impossibility to access it strictly speaking. If crossing the datasets had been strictly prohibited, and hence legally impossible, under German law, the CJEU would have most likely concluded that the dynamic IP addresses held by the website provider were non-personal (i.e., anonymous) data.

Ultimately, the ruling in Breyer restricts the otherwise-expansive scope of the Recital 26 test. The mere existence of ARI in the hands of a third party is not enough in itself to make the data held by a controller personal. Rather, we have to consider whether the data controller has means reasonably likely to be used to access that ARI; and this, in turn, necessitates a contextual evaluation of all the circumstances of a case, from both a logistical standpoint (eg, considering the cost and the amount of time required) and a legal one.

5.8 Pseudonymisation: Attempting a Reconciliation

The GDPR shepherds a new concept into European data protection legislation: that of ‘pseudonymisation’. Article 4(5) GDPR defines pseudonymisation as:

‘the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.’\textsuperscript{82}

The concept of ‘pseudonymisation’ was nowhere to be found in the Data Protection Directive, and appeared for the first time in the GDPR. Although it may seem to play an innocuous, or very limited, role within the legal architecture of the

\textsuperscript{81} See also Purtova, n 45 above, 65 (reaching a similar conclusion).

\textsuperscript{82} General Data Protection Regulation, art 4(5).
GDPR, the notion of ‘pseudonymisation’ is rather problematic. Above all, what should puzzle us is that pseudonymised data (ie, data that have undergone pseudonymisation as per Article 4(5)) are defined as ‘personal’ data a priori under the GDPR. Recital 26 clarifies that:

‘Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person.’

The a priori classification of ‘pseudonymised data’ as ‘personal’ seems inconsistent with the relative small-risk approach to identifiability enshrined in the Recital 26 ‘reasonably likely means’ test and endorsed by the CJEU in Breyer. To better appreciate why, consider the following example.

Company A has conducted a health study, and eventually come up with a dataset containing personal data, including the name of interviewees, their unique ID numbers, and their clinical history. Company A has pseudonymised all the data: it has encrypted any direct identifier in the dataset (eg, name and ID numbers) and stored the decrypting key safely away, putting in place technical and organisational measures to prevent unwarranted re-identification. Company B, a pharmaceutical company, has bought access to the pseudonymised dataset. It has no logistical means of accessing the decrypting key (the additional information needed to re-identify the data subjects in its partial dataset) and it is legally prohibited from even attempting to do so, both under national law and under the terms of its contract with Company A. Is the data held by Company B anonymous or personal?

If we were to simply refer to Recital 26 against the backdrop of the CJEU’s ruling in Breyer, we should conclude that the data held by Company B are anonymous. Logistically, Company B has no ‘means reasonably likely to be used’ to identify the data subjects. In the light of Breyer, the possibility of Company B simply approaching Company A and illicitly asking for the decrypting key is dismissed as an option ‘unlikely’ to be used given its illegality.

However, there might be another way of analysing the example above, yielding a disquietingly different answer. Specifically, we could consider the data held by Company B as ‘data which have undergone pseudonymisation’ and conclude that they are personal, given the apparent categorisation of pseudonymised data as ‘personal’ by default under Recital 26. Oddly enough, this pseudonymisation-focused analysis would contradict that in the paragraph above, and fly in the face of the relative small-risk approach to identifiability spelt out in Recital 26 and supported by the CJEU in Breyer. In light of the foregoing, we have to either at-

---

83 General Data Protection Regulation, recital 26.
84 Ibid.
tempt a reconciliation or acknowledge a disturbing inconsistency in our data protection legislation.

In an attempt to reconcile pseudonymisation with a Breyer-inspired risk-based approach to identifiability, eminent commentators have suggested that the starting point in any ‘identifiability’ analysis should always be the ‘reasonably likely means’ test in Recital 26. That test should determine whether the data in the controller’s hands are anonymous or personal. If the data are found to be personal, they then fall within the remit of the GDPR. If, however, personal data have been protected against the risk of reattribution through pseudonymisation (ie, separating the information enabling attribution and subjecting it to technical and organisational safeguards), it is likely that the controller will more easily satisfy its multiple obligations under the GDPR, including the duty to guarantee data protection ‘by design and default’. But pseudonymisation per se plays no role in determining whether yet-to-be-categorised data should be anonymous or personal. Ultimately, a controller should always engage in a two-step analysis: (1) ask whether its data are personal under Recital 26; (2) and, if so, ask whether they have been pseudonymised within the meaning of Article 4(5). If both questions are answered in the affirmative, then the controller’s data will fall under the GDPR but it will be easier for her to fulfil her other obligations under the GDPR.

The foregoing interpretation of ‘pseudonymisation’ should be endorsed both normatively and textually. Normatively, it avoids an otherwise-inexplicable inconsistency within the GDPR and is therefore to be favoured as a desirable analysis. Textually, it is consistent with the wording of both Article 4(5) and Recital 26. Firstly, Article 4(5) defines ‘pseudonymisation’ as ‘the processing of personal data [emphasis added]’ in such a way as to decrease the risk of reattribution: this presupposes that the data to be pseudonymised are personal. In turn, this implies that there is an earlier step to take to determine whether the data are personal to begin with, and that is most likely the test under Recital 26. Sure enough, if the data are found to be anonymous under Recital 26, Article 4(5) could not apply at all: there could no longer be a processing of ‘personal’ data. Secondly, Recital 26 reiterates that ‘personal data which have undergone pseudonymisation [emphasis added]’ are still personal data. Albeit employing a perplexing wording, the Recital is merely confirming that pseudonymised personal data, even if falling within

86 General Data Protection Regulation art 25(1) (‘the controller shall implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation [emphasis added]’).
Article 4(5), are not ipso facto anonymous. In a sense, it is reiterating that pseudonymisation has no role to play in the ‘identifiability’ analysis itself.

To sum up, whether data are personal or anonymous falls to be determined under the ‘reasonably likely means’ test in Recital 26, bearing in mind the CJEU’s ruling in Breyer. Pseudonymisation is not relevant in the ‘identifiability’ analysis and only steps in at a second stage when the nature of the data has already been determined.

5.9 The Ongoing-Duty Element: a Dynamic Approach to Identifiability

Lastly, data have a dynamic nature. Recital 26 makes this clear when reminding us that:

‘To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments all must be assessed at the time of processing [emphasis added].’

In light of Recital 26, data which have undergone anonymisation may cease to be anonymous if new technologies spring up enabling re-identification. This can be a practical concern for many organisations, and may translate into an ongoing obligation on the part of a controller to make sure that anonymised data remain non-personal. Importantly, this may have also sung the requiescat of the so-called ‘release-and-forget’ model of anonymisation.

5.10 Tying All Together

The foregoing inquiry into the personal-versus-anonymous-data debate may appear a sterile exercise of scholarly elucubration. Yet it has a crucial bearing on any instance of data commercialization, and cannot be ignored if we are to have a

---

87 General Data Protection Regulation, recital 26.
88 For the inherent dynamism of data, see V. Janeček and G. Malgieri, ‘Data Extra Commercium’, in Lohsse, Schulze and Staudenmayer (eds), n 2 above, 12.
89 Ohm, n 69 above (noting the failure of the ‘release-and-forget’ model, whereby a controller assumes data have been anonymised forever and forgets them after having transferred or disclosed them).
sensible discussion about trade in data. By and large, businesses willing to commercialize their data may wish to anonymise them so as to escape the strictures of data protection legislation and smooth over their transactions. As we saw above, however, they may have a hard time slipping out of the fold of Article 4(1) GDPR: the concept of personal data is extremely wide in scope and its boundaries are hard to pin down. Similarly, taking anonymisation lightly may have dire consequences. In 2018, the Danish Data Protection Authority fined Taxa 4xA DKK1.2mln for having failed to anonymise personal information on their customers’ taxi trips. The Danish taxis company had deleted customers’ names from their dataset, but had left in their telephone numbers: these could uniquely identify each passenger and prevented the dataset from being effectively anonymous. In the mistaken belief to be dealing with anonymous data, the company had failed to comply with the provisions of the GDPR (in particular, with the duty not to store data longer than necessary).  

For the purposes of anonymisation, businesses will have to ensure (i) that their data are not about an identified or identifiable data subject; and (ii) that they are not intended nor likely to affect any identified or identifiable data subject. Secondly, and following from the above, they will need to make sure that none of their data relates to any ‘identified or identifiable natural person’. With formerly personal data which have undergone anonymisation, this implies that the risk of someone being re-identified within a dataset must be small to null. Often, businesses will need to go to great lengths to anonymise their data: perhaps, only high-aggregating or computationally-sound techniques (like k-differentiation) can be guaranteed to be sufficient. Encryption, hashing and other pseudonymising techniques replacing one unique identifiers for another will never be enough in themselves to anonymise data. Ultimately, whether data are anonymous is to be determined under the test in Recital 26, considering any means at their controller’s disposal and also those reasonably likely to be used by third parties. If approaching a third party to secure the additional information enabling re-identification is legally prohibited, that shall not count as a means likely to be used in the light of Breyer. Lastly, all businesses holding anonymous data will have an ongoing duty to ensure that their data keep staying non-personal or to otherwise make sure to always have a lawful basis for any further processing.

---


91 Art 29 Working Party, Opinion 05/2014 on Anonymisation Techniques (WP 216) 0829/14/EN.
6 The ‘How’ Problem

We have seen how problematic it is for businesses to anonymise their data. In most common scenarios, however, businesses will be dealing with personal data and will therefore need to pay regard to data protection legislation. Needless to say, the GDPR will apply if a business intends to commercialise its personal data without prior de-identification. Similarly, if a business is commercialising a mixed dataset (with both personal and non-personal data in it), the GDPR will apply to the processing of its personal elements. If the personal and non-personal data in a mixed dataset are ‘inextricably linked’, the GDPR will apply regardless. Beyond, data protection legislation will be a relevant concern for businesses unsure whether their data will count as non-personal and thereby willing to err on the side of caution. Assuming the GDPR applies, let us now turn to the question how it would apply to a commercialisation of data.

6.1 Commercialisation of Data as Processing

Under the GDPR, ‘commercialising’ personal data or ‘trading’ them with another entity will usually be a form of ‘processing’. Crucially, a firm has various ways of commercialising the personal data at its disposal. It may share personal data with a third party (eg, an advertiser or a partner) by giving them access to their datasets. As an example, PayPal reports sharing its consumers’ data (often inclusive of their name, address, email address and phone number) with hundreds of entities across the world, including commercial partners (eg, Apple, DHL), financial product providers (eg, BNP Paribas) and payment processors (eg, American Express). But a firm may also monetise data without directly sharing them with a third party. In particular, a firm (eg, social media platforms) may profile its users and be paid to show them targeted ads based on their preferences, with advertisers buying a ‘window’ to biddable consumers without directly accessing their

---


93 Ibid 9.

94 The ‘Bounty’ scandal is illustrative of the importance of staying compliant with data protection legislation when sharing data with third parties. Bounty, a parenting company, sold highly-sensitive information on millions of mothers and babies in breach of data protection legislation (here, the DPD). In 2019, the ICO fined it £400,000; see M. Murgia, ‘Personal data on new mothers and babies sold to third parties’ Financial Times (London, 19 April 2019).

personal information. This is how Facebook monetises its data. In its privacy policy, the Silicon Valley-based big-tech company reassures its user that ‘No, we don’t sell your information. Instead, based on the information we have, advertisers and other partners pay us to show you personalized ads on the Facebook family of apps and technologies’.96

Finally, a firm may monetise personal data by selling them: not just sharing access to a dataset, but alienating a dataset altogether. We could hyperbolically picture this category in terms of two firms ‘exchanging hard drives full of your data for a suitcase of money’.97 Interestingly, the value that data have (or that they may acquire with the future advances in analytics) is often such as to deter companies from ‘selling’ datasets outright: only sharing access to datasets will often prove a more sensible way of monetising them in the longer run.98 Nevertheless, the whole talk of ‘selling’ data is inevitably contentious. Firms are reluctant to speak of their data-monetisation practices as a form of ‘sale’ for fear of reputational damage. Just like Facebook (as we saw above), Google also makes it clear that ‘we do not sell your personal data to anyone’.99 Conversely, some activist groups, media and politicians often rely on the very words ‘selling’ or ‘sale’ to describe emphatically, and take issue with, big firms’ obscure data-monetisation practices. At the same time, many in the European institutions are uncomfortable with the concept of ‘selling’ data.100 Taking on a fundamental rights perspective, the European Data Protection Board noted that personal information cannot be conceived of as a mere economic asset or commodity in EU Law. Commenting on the proposal for the Digital Content Directive, the EDPS compared the market for personal data to the market for live organs: the empirical fact that there exists a

98 Europe and North America, ‘Your Data Is Shared and Sold...What’s Being Done About It?’ (Knowledge@Wharton) <https://knowledge.wharton.upenn.edu/article/data-shared-sold-whats-done/> accessed 15 December 2020.
100 For a view against recognising the possibility of one ‘selling’ personal data, see the European Data Protection Supervisor (EDPS), Opinion 4/2017 of the European Data Protection Supervisor on the proposal of a Directive on the Supply of Digital Content <https://edps.europa.eu/data-protection/our-work/publications/opinions/contracts-supply-digital-content_en> (‘However, personal data cannot be compared to a price, or money. Personal information is related to a fundamental right and cannot be considered as a commodity.’, [14]).
market for personal data, just like there is one for human live organs, does not mean that it should automatically receive ‘the blessing of legislation’ or that we should thoughtlessly give in and re-conceptualise data as ordinary commodities.  

The stance taken by the EU parliament in no less firm. Albeit accepting the *de facto* economic reality of data commerce with the enactment of Digital Content Directive, the EU Parliament refused to sanction it *de jure*, pushing back against the notion of data as ‘counter-performance’ (advanced by the Commission in its draft proposal on the DCD) and omitting it in the final text of the directive.  

If (i) sharing access to a dataset, (ii) showing targeted ads to consumers based on their personal information, and (iii) selling datasets are the three main instances of ‘commercialisation of data’, they will all be equally caught under the GDPR notion of ‘processing’. Indeed, the word ‘processing’ broadly refers to:

‘...any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.’

Plainly, ‘sharing access’ to a dataset is a way of ‘making [data] available’; ‘using’ consumers’ personal data to profile them and show them targeted ads is trivially a way of ‘using’ data; and ‘selling’ data outright is simply another form of ‘disclosure’ or of ‘making [data] available’.

A few more terminological issues are worth addressing here. A data ‘controller’ is any natural or legal person determining the purposes of the processing; on the other hand, a data ‘processor’ is someone (other than a data controller’s employee) that processes data on behalf of the data controller. A commercialisation of data is likely to be a form of controller-to-controller transaction. As such, it is subject to a different set of regulations than a controller-to-processor

---

101 Ibid [17].  
103 ‘Data trade’ and ‘data commerce’ are considered synonymous with ‘commercialisation of data’ (although it may be argued that ‘data trade’ should cover instances (i) (sharing access to datasets) and (iii) (selling datasets), but not (ii) (since data do not move here at all)).  
104 General Data Protection Regulation, art 4(2).  
105 General Data Protection Regulation, art 4(7).  
106 General Data Protection Regulation, art 4(8).
transfer, the latter falling beyond the scope of this paper.\textsuperscript{107} We shall now investigate the GDPR principles that would govern commercialisation of data within the EU.\textsuperscript{108}

### 6.2 Fairness, Lawfulness and Transparency

Firstly, personal data must be processed ‘lawfully, fairly and in a transparent manner’ under the GDPR.\textsuperscript{109} Among other things, this translates into an inescapable imperative: to provide information to data subjects. In the context of commercialisation, the data-trader may have informed its data subjects already about the likelihood of a future commercialisation. If this is not the case, data-traders will have to notify them of the intended commercialisation. Information must be provided ‘in a concise, transparent, intelligible and easily accessible form, using clear and plain language’\textsuperscript{110}.

Under the GDPR, controllers are under a duty to inform data subjects both when their data are directly collected, and when they are \textit{indirectly} collected.\textsuperscript{111} Therefore, all data-recipients will also be under a duty to notify data subjects of the relevant commercialisation under Article 14. They will need to do so within one of the following three timeframes: (i) within a reasonable period, but within one month at the latest; (ii) when the data bought are to be used to communicate with the data subject, at time of the first communication at the latest; (iii) if the parties intend to further disclose the data to another recipient, at the time the data are first disclosed at the latest.\textsuperscript{112} In light of the above, the data-trader will ultimately have to assess what information they have provided already. Equally, any data-recipient will need to consider when and how to inform the data subjects under Article 14(3).

\textsuperscript{107} This is dealt with under General Data Protection Regulation, ch 4.
\textsuperscript{108} The scope of transfers outside are beyond the scope of this paper. Briefly put, cannot be transferred outside the EEA to unless that country or territory guarantees an adequate level of protection for data subjects or another legitimate basis for the transfer exists. See General Data Protection Regulation, art 44. The buyer will also have to ensure that any transfer outside the EEA has been executed in compliance with applicable data protection legislation protection.
\textsuperscript{109} General Data Protection Regulation, art 5(1)(a).
\textsuperscript{110} General Data Protection Regulation, art 12(1).
\textsuperscript{111} General Data Protection Regulation, art 13–14.
\textsuperscript{112} General Data Protection Regulation, art 14(3).
6.3 Lawful Basis for Processing (Data-Trader’s Side)

Under the GDPR, the data-trader will need a lawful basis for commercialising its data.113 The following three bases are most likely to be successfully invoked: ‘consent’ (Article 6(1)(a) GDPR); ‘contractual necessity’ (Article 6(1)(b) GDPR); and ‘legitimate interests’ (Article 6(1)(f) GDPR).

6.3.1 Consent

The controller can invoke consent as a legal basis for processing if the data subject has consented ‘to the processing of his or her personal data for one or more specific purposes’.114 Historically, a controller’s usual way of securing consent was by concluding an agreement with the data subject and incorporating consent into its wording (eg, a privacy policy requiring users to agree to its terms before they could access a digital service).115 However, the GDPR makes consent harder to rely on. Under Article 7, the data-trader should ensure that the data subjects have consented unambiguously to the commercialization of their data; that they were fully informed about it; that their consent remains valid and has not been revoked if the data controller is relying on a consent previously secured; and that consent is specific and freely given, eg, that the data subjects have the realistic option not to consent without suffering disproportionate detriment.116

This legal basis will often be the safest option to fall back on for firms willing to err on the side of caution.117 Unfortunately, this has undesired side effects: in particular, the notion of ‘consent’ is stretched to the breaking point in the digital environment. In this regard, Margaret Radin has influentially argued that ‘massively-distributed boilerplate’ (also known as ‘adhesion contracts’) are upending traditional categories of contract and consent.118 In particular, we often rely on the

---

113 General Data Protection Regulation, art 6.
114 General Data Protection Regulation, art 6(1)(a).
116 General Data Protection Regulation, art 7; General Data Protection Regulation, recital 32.
‘objective theory of contract’ in contract law in asking what a reasonable observer would take the parties to a contract to have intended.\textsuperscript{119} The parties’ \textit{subjective} state of mind is often marginal in the analysis. On Radin’s account, the objective theory of contract has been distorted, and its purpose perverted, to fit boilerplate into legal doctrine. This is particularly evident in the digital space, where it is as hard to navigate the ubiquitous fine print as it is easy to consent to it. Consider, for instance, a consumer clicking ‘I agree’ on a tiny digital box, with a link redirecting her to the fine print she is supposedly agreeing to.\textsuperscript{120} Under a misplaced objective theory of contract, we take all such clicks ‘metaphorically’, as conveying genuine acceptance of contractual terms just like someone’s signature on a bespoke contractual document.\textsuperscript{121} On Radin’s view, this is like turning a deaf ear to reality. There is abundance of empirical evidence showing that consumers are unlikely to read boilerplate or even simply to understand it.\textsuperscript{122} If it is socially known that consumers rarely, if ever, engage with the fine print, taking a consumer’s consent at face value is no longer a warranted practice in the digital space. It is not reasonable for a firm to infer from a consumer’s effortless click that they are \textit{genuinely} consenting to a set of terms, nor is a ‘reasonable observer’ justified in taking a consumer’s click to imply \textit{consensus ad idem}.\textsuperscript{123} This practice is making ‘liars of us all’, as Radin stingingly remarks.\textsuperscript{124} It follows that massively-distributed boilerplate cannot accommodate such an objective notion of consent as was intended under traditional categories of contract law: contract and objective consent are simply not the right \textit{lenses} to see through what goes on in the digital environment. Doing away with a purely contractual framework, and shifting to a consumer-protection regulatory regime, will often be a fairer and more socially-honest option.\textsuperscript{125}

The GDPR \textit{partly} addresses the concerns raised by Radin on the inadequacy of an objective theory of consent. By requiring the data subject’s consent to be ‘freely given’,\textsuperscript{126} and requiring it to be ‘explicit’ insofar as sensitive data are concerned,\textsuperscript{127} the GDPR seems to favour an autonomy-centred \textit{subjective} notion of consent (just

\begin{itemize}
  \item \textsuperscript{119} Radin (2017), n 118 above, 518.
  \item \textsuperscript{120} Ibid 519.
  \item \textsuperscript{121} Ibid 520.
  \item \textsuperscript{123} Radin (2017), n 118 above, 520.
  \item \textsuperscript{124} Radin (2017), n 118 above, 519.
  \item \textsuperscript{125} See Radin (2017), n 118 above, 533 (considering ‘generalised regulation’ as being a better option).
  \item \textsuperscript{126} General Data Protection Regulation, art 7(4).
  \item \textsuperscript{127} General Data Protection Regulation, art 9(2)(a).
\end{itemize}
as we would require subjective ‘informed consent’ in, say, medical settings).\textsuperscript{128} It would also be too uncharitable of us to say that the GDPR sees data transactions \textit{only} through contractual lenses under Article 6(1)(a). Even if a consumer agrees to the processing of her data, she may withdraw her consent at any point further down the line and the data controller is bound to the multiple obligations imposed by the GDPR all throughout.\textsuperscript{129} In a sense, a consumer can agree to the processing of their data, but cannot agree to transfer their \textit{right} to personal data away.\textsuperscript{130}

Still, the GDPR does not entirely solve the crux of the problem either. If a data privacy note is framed in a supposedly clear and intelligible language, consent by clicking ‘I agree’ on a digital inbox may well be sufficient.\textsuperscript{131} As an example, Facebook’s data privacy notice suggests that consent is one (and, arguably, the main one) of the legal bases that the company relies on to ensure compliance with the GDPR.\textsuperscript{132} Yet it is doubtful that most Facebook users have perused the terms of the privacy policy they are taken to have agreed to. Said otherwise, the dangers of a vacuous residually-objective notion of consent are not entirely averted by the GDPR. A risk remains that, even \textit{when} the GDPR applies under Article 6(1)(a), the threat to privacy still stands.

\subsection*{6.3.2 Contract}

The controller can lawfully process data if ‘the processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract’.\textsuperscript{133} To clarify when processing is deemed necessary for the performance of a contract, the European Data Protection Board endorsed the Article 29 Working Party’s guidance on an equivalent provision in the Data Protection Directive;\textsuperscript{134} namely that:

\begin{itemize}
\item \textsuperscript{128} Radin (2014), n 118 above, 89.
\item \textsuperscript{129} General Data Protection Regulation, art 7(3).
\item \textsuperscript{130} European Data Protection Board, ‘Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects’, [54].
\item \textsuperscript{131} Insofar as the matters that the consumer is being asked to agree to are kept separate; see General Data Protection Regulation, art 7(2).
\item \textsuperscript{133} General Data Protection Regulation, art 6(1)(b).
\item \textsuperscript{134} European Data Protection Board, ‘Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects’.
\end{itemize}
‘[the provision] must be interpreted strictly and does not cover situations where the processing is not genuinely necessary for the performance of a contract, but rather unilaterally imposed on the data subject by the controller’.135

In other words, the processing must be ‘objectively necessary’ to achieve the purpose of the contract in such a way that the contractual service could not be delivered without the processing.136 To give an example, if a customer wishes to make a purchase on an online store and pay with their credit card, the retailer will have to process the credit card details and billing address to deliver on the customer’s request. Equally, the retailer may have to process the customer’s residential address if they have requested home delivery. Conversely, if the customer has requested the purchased items to be delivered at a pick-up location, it will be unnecessary to process the customer’s residential address for the contractual service to be performed.137 In the labour context, for instance, processing an employee’s bank account details can be necessary to pay salaries and perform the employment contract.138

Often, a contractual agreement concluded with a data subject may refer to additional processing activities – eg, profiling, targeting and advertising. However, the data controller will be unable to argue that they form part of the contractual purpose of the agreement: whether something is genuinely necessary for the performance of the contract depends on the mutually understood purpose of the agreement and therefore also on the data subject’s understanding of what the contract is about.139 Someone buying an item of clothing online does not expect the contract entered into with the retailer to be about a profiling of their tastes and lifestyle choices. Rather, they expect the contract to be about delivering goods in exchange for monetary value.140 The retailer may have referred to these extra processing activities in the small print of the contract. However, this does not make them any more ‘necessary’ for the performance of the contract than they genuinely are.141

Alternatively, a data controller can show that the processing is necessary ‘in order to take steps at the request of the data subject’ prior to entering into a con-

136 European Data Protection Board, n 130 above, 8.
137 Ibid 10–11.
138 Art 29 Working Party, n 135 above, 16.
139 European Data Protection Board, n 130 above, 10.
140 Ibid 11.
141 Ibid.
tract [emphasis added]'\textsuperscript{142} Often, a preliminary data processing may be necessary to facilitate the parties’ future entry into the contract. As made clear by use of the phrase ‘at the request of data subject’, the provision is not intended to cover instances of processing carried out at the request of third party or on the initiative of the data controller.\textsuperscript{143} Rather, it may be referring, for instance, to a data subject providing their address or postal code to inquire whether a given service is available in their neighbourhood. Processing the address or postal code would be necessary to take steps at the request of the data subject prior to the entering into the contract in that case.\textsuperscript{144}

It is worth noting here that there is a clear interplay between Article 6(1)(b) GDPR and Article 3 of the Digital Content Directive (hereinafter, ‘DCD’).\textsuperscript{145} Enacted in 2019, the DCD fills a lacuna in the existing European consumer-rights framework by extending legal protection to contracts for the supply of digital content and digital services. Under Article 3, the DCD applies to ‘any contract where the trader supplies or undertakes to supply digital content or a digital service to the consumer and the consumer pays or undertakes to pay a price’, but also ‘where the consumer provides or undertakes to provide personal data [emphasis added]’.\textsuperscript{146} The European legislator here acknowledges that online goods or services may be paid for with personal data, albeit taking issue with the notion of data as ‘counter-performance’ (advanced by the Commission in its draft proposal on the DCD) and expunging it from the final text of the directive.\textsuperscript{147} Importantly for our purposes, the DCD does not apply ‘where the personal data provided by the consumer are exclusively processed by the trader for the purpose of supplying the digital content or digital service in accordance with this Directive’.\textsuperscript{148} All things being equal, this may well suggest that successfully relying on Article 6(1)(b) GDPR as a legal basis for the processing should imply that the DCD does not bite. Consider, for instance, a company processing data to discharge a duty under

\textsuperscript{142} General Data Protection Regulation, art 6(1)(b).
\textsuperscript{143} European Data Protection Board, n 130 above, 13.
\textsuperscript{144} Ibid.
\textsuperscript{146} Digital Content Directive, n 145 above, art 3.
\textsuperscript{148} Ibid.
a contract for the supply of online services: an online hotel search engine processing a customer’s personal data to show her some nearby hotels, under a contract for the online display of hospitality services based on some pre-defined search criteria. The processing of data is here allowed as a ‘contractual necessity’ under Article 6(1)(b) GDPR, and it follows too that DCD should not apply under Article 3(1).

This overlap has an unintended consequence: the difficulties bedevilling the interpretation of Article 6(1)(b) GDPR will also affect Article 3 DCD. For example, it is unclear if data processed for the purposes of online behavioural advertising (that is, for the purposes of monetising them through targeted advertising) are necessarily outside the scope of Article 6(1)(b). It is indeed hard to argue that monetising A’s data for the purposes of online advertising is necessary for the purposes of a contract entered into with A, but it may not be impossible to ever make that case. The European Data Protection Board indeed confirms that that ‘as a general rule, processing of personal data for behavioural advertising is not necessary for the performance of a contract for online services [emphasis added]’. Still, one cannot make the case that monetising data through targeted advertising is necessary to fund the online service (e.g., Facebook) and that, by implication, processing data is necessary for the controller to discharge her contractual obligation under the contract.

In light of the above, a data controller intending to commercialise their data will have to prove that the commercialisation of data is genuinely necessary for the performance of the contract entered into with the data subject. A mere reference to the possibility of a future commercialisation in the small print of an agreement will not make it any more necessary to the performance of the contract than it genuinely is. Alternatively, data-traders will have to show that the relevant commercialisation is necessary for the taking of steps at the data subject’s request prior to the entering into the contract. In most cases, they will be hard put to do so.

### 6.3.3 Legitimate Interests

The controller can process data lawfully if ‘the processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party,
except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data’.  

Firstly, the processing must be ‘necessary’ to serve the interest at stake. Interpreting a similarly-worded provision in the Data Protection Directive, the Article 29 Working Party suggested that controllers should ‘consider whether there are other less invasive means to reach the identified purpose of the processing and serve the legitimate interest of the data controller’. In the context of data commercialization, it may be argued that the immediate interest being pursued by the data controller is a monetary profit. Accordingly, a case could always be made that there are other less-intrusive means to secure a profit: eg, content providers could differentiate their service offer and provide paid-for premium services. Secondly, the interest pursued must be ‘legitimate’. There seems to be broader room here to argue that an interest qualifies as legitimate. If lawful, sufficiently clear, and representing a real and present interest, a monetary profit may as well be legitimate. Thirdly and lastly, a controller’s legitimate interest should be balanced against the impact on the data subjects. For instance, children’s interest in data protection will often be paramount and outweigh any commercial profit that a controller could derive from selling their data.

Tech firms often rely on this legal basis to address their data-commercialisation practices, although they may usually do so in conjunction with consent (Article 6(1)(a)) and contractual necessity (Article 6(1)(b)). However, firms are unlikely to rely on ‘legitimate interests’ as a basis for more invasive data-monetisation practices, like intrusive profiling and targeting carried out for advertising purposes. Indeed, the Article 29 Working Party considered that it would be hard to satisfy this legal basis with ‘intrusive profiling and tracking practices for marketing or advertising purposes, for example those that involve tracking individuals across multiple websites, locations, devices, services or data-brokering’.

It is more likely that firms will rely on this legal basis to address less intrusive processing for which no consent has been previously secured. As an example, Google reports relying on ‘legitimate interests’ as a legal basis ‘when using personal data

152 General Data Protection Regulation, art 6(1)(f).
153 Art 29 Working Party, n 135 above, 55.
154 As observed by Janeček and Malgieri, n 88 above, 12.
155 Art 29 Working Party, n 91 above, 24, 49.
156 General Data Protection Regulation, recital 38; see also art 39 Working Party, Opinion 2/2009 on the protection of children’s personal data (General Guidelines and the special case of schools) (WP160).
for activities such as serving contextual ads, ads reporting and to combat fraud and abuse. Specifically, this is the basis that Google relies on when showing users non-personalised ads, using coarse (e.g., city-level) personal information without users’ prior consent. Facebook reports relying on ‘legitimate interests’ in its data policy too. It invokes ‘legitimate interests’ as a legal basis, among other things, (i) ‘for providing measurement, analytics [...]’ to their partners (the legitimate interest here being to ‘to provide accurate and reliable reporting to our advertisers, developers [...]’); as a basis (ii) ‘for providing marketing communications’ to the user (the interest here being to ‘to promote Facebook Company Products’); as a basis (iii) ‘to research and innovate for social good’ (the interest here being ‘to further the state-of-the-art or academic understanding on important social issues); and finally, as a legal basis (iv) ‘to share information with others including law enforcement and to respond to legal requests’ (the interest here being, among other things, ‘to prevent and address fraud’). Insofar as more intrusive data-monetisation practices are concerned, firms will usually seek to rely on consent under Article 6(1)(a). As a vivid example, Facebook explicitly invokes consent as a basis for using users’ personal data to show them targeted ads, with their terms stating that ‘by using our Products, you agree that we can show you ads that we think will be relevant to you and your interests. We use your personal data to help determine which ads to show you’.

6.4 Sensitive Data

Neither the ‘legitimate interests’ nor the ‘contractual necessity’ grounds will be available if the data being traded are not just personal, but also sensitive. Sensitive data are defined in Article 9(1) GDPR as:

‘personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation’.

---

159 Ibid.
160 Art 29 Working Party, n 157 above.
162 General Data Protection Regulation, art 9(2).
163 General Data Protection Regulation, art 9(1).
The GDPR prohibits the processing of sensitive data, save in a number of circumstances listed under Article 9. Arguably, sensitive data could only be commercialised in two cases: (1) when ‘explicit consent to the processing of those personal data for one or more specified purposes’ is secured; or (2) when the processing relates to ‘[sensitive] data which are manifestly made public by the data subject’. However, this latter instance is unlikely to come up in the context of commercialization of data: there would be no need to pay for data that have been already made available publicly. Therefore, a data-provider wishing to commercialise sensitive data will often need to rely on explicit consent. It has been observed that ‘explicit consent’ can be compared with regular ‘consent’. However, while the GDPR spells out the prerequisites of regular consent, it fails to articulate whether there should be any distinction between ‘explicit consent’ and ‘regular’ consent, and, if so, what that distinction would amount to. It follows that businesses wishing to trade sensitive data will confront serious difficulties in practice: importantly, they will be in doubt as to what steps they ought to take in order to secure ‘explicit’ consent, and whether they need to do more than what they would ordinarily do to obtain ‘regular’ consent.

Lastly, there is another point worth considering. When discussing personal and non-personal data, we observed that the conceptual boundaries separating the two are often blurry and ill-defined. We also saw that the nature of data is inherently dynamic: anonymised data may go back to being personal if new technologies emerge enabling re-identification. A similar dynamism applies to sensitive data. With artificial intelligence, machine learning and big data analytics, we may be able to infer sensitive data from what looks like non-sensitive information. For instance, with a capillary harvesting of data and the further development of predictive analytics, we may be able to statistically infer a data subject’s health conditions from nothing but their daily geo-location data points. When commercializing data, both the data-trader must therefore be responsive to the dynamic nature of data not just in terms of identifiability, but also of sensitivity. After all, whether the data are sensitive will affect both whether the commer-

164 General Data Protection Regulation, art 9(2)(a).
165 General Data Protection Regulation, art 9(2)(e).
cialisation can take place (depending on whether a lawful basis for processing under Article 9 GDPR is available) and how it can take place.

### 6.5 Lawful Basis for Processing (Data-Recipient’s Side)

Simply put, the data-recipient, if any, must be satisfied that the personal data were lawfully collected by the data-provider to begin with. Furthermore, it must also be satisfied that there is an ongoing valid basis for processing them.

### 6.6 Other Data Protection Principles

Processing data ‘lawfully, fairly and in a transparent manner’ may not be enough to meet a controller’s obligations under data protection law. The parties to the data commerce may need to pay regard to the remaining five data protection principles under the GDPR: purpose limitation; data minimisation; accuracy; storage limitation; and data integrity and confidentiality. They will also need to meet the overarching accountability principle demanding ongoing compliance with the data protection system as a whole. Paying regard to these principles may require the data-recipient to investigate the data-provider’s approach to various areas, including: (i) whether the use of the database was limited to the purpose for which the data were originally collected (purpose limitation); (ii) whether data are limited to what is necessary to achieve the data controller’s stated purpose (data minimisation); (iii) whether data are accurate and up-to-date (accuracy); (iv) what data are being retained and how (storage limitation); and (v) whether data security is guaranteed (integrity and confidentiality).

---

168 If the data commerce happens without the counterparty ever accessing the dataset (eg, with an advertiser paying Facebook to show targeted ads to a range of users), there may technically be no data recipient at all.


170 General Data Protection Regulation, art 5(1)(a).

171 General Data Protection Regulation, art 5(1)(b).

172 General Data Protection Regulation, art 5(1)(c).

173 General Data Protection Regulation, art 5(1)(d).

174 General Data Protection Regulation, art 5(1)(e).

175 General Data Protection Regulation, art 5(1)(f).

176 General Data Protection Regulation, art 5(2).
Although the matter falls strictly beyond the scope of this paper, it is worth noting in passing that legal frameworks other than data protection law may be just as crucial to data commerce. For instance, a refusal to trade ‘unique and non-substitutable data owned by a firm in a dominant market position’ may raise thorny issues from a competition law perspective. On the front of intellectual property law, copyright might subsist in databases ‘which, by reason of the selection or arrangement of their contents, constitute the author’s own intellectual creation’; and a sui generis right may exist to protect databases that result from a substantial investment in the ‘obtaining, verification or presentation of the contents’ of a database and irrespective of their originality. Valuable commercial data may also be protected as ‘trade secrets’. Any recipient of data will want to clear its new datasets of any of these pre-existing rights. Lastly, the European Convention of Human Rights is another legal framework to keep in mind when thinking around data commerce. In particular, Article 10 ECHR enshrines the fundamental right to freedom of ‘information’. If ‘information’ is read as extending, by implication, to data, Article 10 would militate in favour of the free flow of data, be it commercial or otherwise.

7 Conclusion

Data are a new goldmine in today’s hyper-connected world. The Internet of Things, data-mining technologies, predictive analytics, cloud computing and data virtualization: these and many more advances in technology have made it possible to collect huge volumes of data, analyse them at unprecedented speed, and build insights on people and their behaviours. Against this background, it is no surprise that European businesses will often wish to commercialise their data and capitalise on their monetary potential. To do so undisturbedly, they will need to stay compliant with data protection legislation. In the present paper, we have

177 Duch-Brown, Martens and Mueller-Langer, n 49 above, 20 (also nothing that there is no extensive case law in this respect).
179 Ibid, art 7(1).
hopefully shown that commercialising data in compliance with data protection law in Europe can be more problematic than it may appear at first.

In particular, there will be two significant hurdles in the way of abiding by data protection law. Firstly, it will often be unclear if the GDPR will bite to begin with (what we have called the ‘if problem’). Indeed, the GDPR applies to personal data only. As we have seen, however, the notion of what constitutes ‘personal data’ remains elusive and might prove extremely wide in scope. Anonymising personal data to de-personalise them and escape the GDPR regime may be very hard, if not practically impossible, in many circumstances. And even where formerly-personal data have been ex hypothesi anonymised, they may cease to be anonymous if new technologies arise enabling re-identification. Secondly, even where the GDPR applies, it will often be difficult to predict how it applies (what we have called the ‘how problem’). It is unclear what lawful basis for processing can be invoked to trade data; nor is it always easy to demarcate sensitive data from non-sensitive ones: their dividing line is as much shifty as it is essential in determining how a business should go about commercialising their data. Ultimately, the ‘Protean’ nature of data and the conceptual elusiveness of many categories in our legal discourse may make it hard for businesses to abide by data protection law in practice. Eventually, businesses and data protection legislation may well end up being more foes than friends.

182 The adjective ‘Protean’ comes from the shape-shifting sea god Proteus in Hellenic mythology. Like Proteus, data can change shape too. Depending on technological advances and on new information becoming available, non-sensitive data may become sensitive again; and personal information can be extracted from once-anonymous data.