The Proposed Data Protection Regulation and Its Potential Impact on Social Sciences Research in the UK

*Leslie Stevens**

This article critically assesses the potential impact of the proposed Data Protection Regulation on the undertaking of social sciences research in the UK, providing practical analysis from the perspective of research involving administrative data. This assessment reveals how changes to the key concepts of anonymisation, personal data and lawfulness may impact upon social sciences research. The approach taken to the regulation of personal versus anonymised data in the proposed Regulation represents a disproportionate and de-contextualised response to the risks involved in undertaking social sciences research that may create disincentives for investing in privacy protective mechanisms. It is positive that there is explicit recognition of research as a legitimate form of data processing. However, negative implications will arise from the introduction of pseudonymous data as a subset of personal data, without proportionate consideration of varying processing contexts and factors surrounding de-identification and specifically, the strict security measures taken to prohibit de-identification in the research context.

I. Introduction

Three years have passed since the European Commission introduced the proposed General Data Protection Regulation ('GDPR')[1] – three drafts have since been released providing an opportunity to approximate what will be the final outcome of trilogue negotiations (expected by the end of 2015).[2] Even after the enactment of the GDPR, there will be a transition phase from current Member State law based on the Data Protection Directive 95/46/EC (DPD).[3] Given the drastic changes introduced in the GDPR from the

* Research Fellow, Administrative Research Centre Scotland and PhD Candidate Mason Institute, University of Edinburgh School of Law. The author would like to thank Professor Graeme Laurie for his contributions to and help in revising initial drafts of this article, as well as to Judith Rauhofer and to peer reviewers for their helpful comments. This work was supported by the Economic and Social Research Council grant number ES/L007487/1 (Administrative Data Research Centre - Scotland). All websites were accessed on 30 July 2015.

1 'Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data' [2012] COM (2012) 11 final. (Hereinafter 'GDPR') To distinguish between the Commission's draft, Parliament's draft and the Council's most recent general approach, references will be made to the 'Tripartite Version' of the GDPR released in 2015 that provides all three in tandem. It will be cited to as e.g. GDPR, Tripartite Version, Parliament, art 3; if no reference is being made to a particular draft, only the relevant article will be cited to. See: GDPR, Tripartite Version -<http://amberhawk.typepad.com/files/eu-council-dp-reg-4column-2015_april.pdf>. The Council's recently adopted general approach is available here: Council of the European Union, 'Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) - Preparation of a General Approach' <http://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf>.

2 'Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data' [2012] COM (2012) 11 final; European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011, C7-0025/2012) (2014) <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P7-TA-2014-0212+0+DOC+XML+V0//EN> ; Council of the European Union, 'Proposal for a Regulation of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation) - Preparation of a General Approach' <http://data.consilium.europa.eu/doc/document/ST-9565-2015-INIT/en/pdf>.

3 Directive 95/46/EC of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data. (Hereinafter 'DPD')

DPD's current legal position it is imperative that those involved in data processing within the EU understand these changes and how it is likely to impact their interactions with data subjects. This article critically assesses the potential impact of the GDPR on social sciences research in the UK from the particular perspective of research involving *administrative data*, which are data originally collected by public sector organisations in their administration of local, regional and national government. Social sciences research carried out under the auspices of the UK's Administrative Data Research Network ('ADRN') will specifically be considered.[4] The resulting analysis demonstrates how key concepts in the GDPR diverge from the current legal position under the DPD, which presents both positive and potentially negative implications for the undertaking of social sciences research in the UK.

1. Potential implications for social sciences research involving administrative data

In 2013, the UK Government's Economic and Social Research Council ('ESRC') invested over £34 million to establish four *administrative data research centres* ('ADRC') that formed the UK's *Administrative Data Research Network* ('ADRN').[5] The ADRN represents 'a UK-wide partnership between universities, government departments and agencies, national statistics authorities, the third sector, funders and researchers.'[6] Through each ADRC, in Northern Ireland, Scotland, Wales and England, the ADRN assists 'accredited researchers [to] carry out social and economic research using linked, de-identified administrative data – information which is routinely collected by government organisations.'[7] Access to de-identified data is provided under strict governance measures to safeguard individuals' rights and interests in their data. These safeguards include a robust process of de-identification involving 'Trusted Third Parties' so that researchers never have access to identifiable data;[8] access that is provided only in secure and monitored settings;[9] and procedural safeguards, including an approvals panel to ensure each application for access represents research which is scientifically sound, non-commercial, ethical and demonstrates a clear potential public benefit.[10]

The social sciences research carried out under the auspices of the ADRN relies upon access to de-identified administrative data, which are originally collected by public sector organisations in their administration of government and delivery of public services. The reuse of data originally collected for purposes other than research is commonplace across the social sciences but also in context with health and biomedical research. In terms of social sciences research and administrative data, the latter encapsulate a wide range of information, including data on births and marriages, income level, social welfare benefits, individuals' housing status, education levels, incidence of crime, child welfare, etc.[11] The GDPR has the potential to disrupt the already robust legal, technical and organisational arrangements (referenced above) which provide safeguarded access to administrative data for research which serves the public interest[12] in 'promoting and improving economic growth, personal and social well-being, and maximising the interests of current and future generations of citizens in the UK.'[13]

However, there are both positive *and* negative implications for research to be drawn from the introduction of the GDPR, in consideration of the general direction taken in all three drafts regarding the concepts of anonymisation, personal data and lawfulness. Understanding the wider implications of

4 Administrative Data Research Network, 'About Us' (2015) <http://adrn.ac.uk/about>.

5 Economic and Social Research Council, 'The Big Data Family Is Born - David Willetts MP Announces the ESRC Big Data Network' (10 October 2013) <http://www.esrc.ac.uk/news-and-events/press-releases/28673/the-big-data-family-is-borndavid-willetts-mp-announces-the-esrc-big-data-network.aspx>.

6 Administrative Data Research Network, 'About Us' (n 4).

7 Administrative Data Research Network, 'About Us' (n 4) (emphasis added).

8 ADRN, 'Trusted Third Parties' <http://adrn.ac.uk/protecting-privacy/de-identified-data/trusted-third-parties>.

9 ADRN, 'Secure Environment' <http://adrn.ac.uk/protecting-privacy/secure-environment>.

10 ADRN, 'Protecting Privacy: Project Approval' (2015) <http://adrn.ac.uk/protecting-privacy/project-approval>.

11 ADRN, 'Administrative Data' <http://adrn.ac.uk/admin-data>.

12 For instance research that addresses issues in social mobility improves understanding of access and support needs for social care of the elderly, informs policies designed to tackle poverty, provides evidence on issues affecting social care for children, etc. See: Administrative Data Taskforce, 'The UK Administrative Data Research Network: Improving Access for Research and Policy' (ESRC, MRC and Wellcome Trust 2012) 1 <http://www.esrc.ac.uk/_images/ADT-Improving-Access-for-Research-and-Policy_tcm8-24462.pdf>.

13 Graeme Laurie and Leslie Stevens, 'The Administrative Data Research Centre Scotland: A Scoping Report on the Legal & Ethical Issues Arising from Access & Linkage of Administrative Data' [2014] Edinburgh School of Law Research Paper No. 2014/35 <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2487971>.

www.lexxion.eu

these particular legislative changes is crucial for the social sciences research community given that the GDPR is now in the final phase of trilogue negotiations.[14] While all three drafts of the GDPR have been released on the basis that 'nothing is agreed until everything is agreed'[15], positions taken on these fundamental concepts reveal major shifts from the current legal position that will undoubtedly impact research. As of June 2015, a general approach to the GDPR has been agreed by the Council of Ministers of the European Union ('the Council'), which provides a compromise position between the European Commission ('the Commission') and the European Parliament's ('the Parliament') drafts of the GDPR.[16]

2. Outline

The article begins in Section II with an overview of the current legal arrangements that govern and make possible the use of administrative data in social sciences research in the UK, from the perspective of the ADRN. Subsection 1 then contrasts existing conceptions of anonymisation and identifiability under the DPD, to the changes introduced by the GDPR that undoubtedly recognise, more fully, the limits to anonymisation. The potential implications of this new standard of anonymisation and identifiability for social sciences research are explored in detail. This analysis leads to Subsection 2, which provides an assessment of the related changes made to the definition of personal data in the GDPR, focussing on the formal addition of *pseudonymous* data as a subset, which has serious implications for research. Subsection 3 reflects on the positive addition of a legal ground for processing on the basis of conducting research, without the need to rely on an additional ground such as consent. Section III concludes the article with more explicit consideration of the potential benefits to social sciences research from the increased recognition of research as a valuable form of processing in itself, while offering an assessment of the potentially negative impact of introducing the new 'sub-category' of personal data in 'pseudonymous data' without allowance for more nuanced and proportionate consideration of the issues at stake for individual data subjects' and wider societal interests.

II. Legal basis for processing administrative data for research

Prior to the establishment of the ADRN, the UK's Administrative Data Taskforce Report in 2012 ('The UK Administrative Data Research Network: Improving Access for Research and Policy') considered the untapped value of administrative data resources in the UK and supported the more efficient use of these under-utilised resources for the public good:

> National administrative data collections held by government departments or agencies that relate to persons and/or organisations have the potential to provide a robust UK-wide evidence base that would contribute a rich new resource for research and policy making and evaluation. Improving access to and linkage between administrative datasets for research and statistical purposes would have demonstrable effects on economic growth and would help us respond more effectively to challenges related to the health and well-being of people.[17]

Access to relevant and quality data is crucial to the undertaking of scientifically sound and ethically robust social sciences research. However, this need must always be tempered by considerations for the rights and interests of data subjects. In context of administrative data, and thus data originally collected by public sector organisations in the course of administering government, the protection of individuals' rights and interests in their data are of the utmost importance, not least because 'Individuals are often under a statutory obligation to provide personal data to the relevant authorities meaning that they lack a freedom of choice: they have to disclose personal data.'[18] The often-obligatory nature of the collection of administrative data provides a persuasive argument in favour of strong data protection in context

14 European Council, Council of the European Union, 'Data Protection: Council Agrees on a General Approach' (15 June 2015) <http://www.consilium.europa.eu/en/press/press-releases/2015/06/15-jha-data-protection/>.

15 'Data Protection: Council Agrees on General Principles and the "One Stop Shop" Mechanism' (*Consilium*) <http://www.consilium.europa.eu/en/press/press-releases/2015/03/13-data-protection-council-agrees-general-principles-and-one-stop-shop-mechanism/>.

16 European Council, Council of the European Union (n 15).

17 Administrative Data Taskforce (n 12).

18 Peter Blume, 'The Public Sector and the Forthcoming EU Data Protection Regulation' 1 European Data Protection Law Review 32, 33 (emphasis added).

of the reuse of public sector data for research.[19] This is why under the auspices of the ADRN, only approved and accredited social sciences researchers can obtain access to de-identified administrative data through a robust process facilitated by Trusted Third Parties.[20] This process provides robust safeguards against re-identification of individuals, either directly or indirectly.[21] Ensuring data are neither directly nor indirectly identifiable is crucial to the protection of individuals' rights and interests whereby effective anonymisation means that the DPD will not apply to such processing.

1. The current standard of anonymisation

Access to previously existing data sets (such as administrative data) is often provided to researchers on the basis that such data are effectively anonymised prior to access being granted. For data to be considered effectively anonymised and thus outwith the scope of the DPD, data must be 'rendered anonymous in such a way that the data subject is no longer identifiable' and that identification is 'no longer possible'.[22] No further guidance is provided in the DPD as to the required standard of anonymisation but the Article 29 Working Party provides guidance that Recital 26 requires 'that data must be processed in such a way that it can no longer be used to identify a natural person by using "all the means likely reasonably to be used" by either the controller or a third party' (this references the part of Recital 26 relating to identifiability and determining what data are identifiable and thus personal data).[23] The Working Party further stipulates that anonymisation should be irreversible,[24] which is an arguably *stricter* standard than compared to that proposed by the UK's Information Commissioner's Office (ICO) in their 'Anonymisation Code of Practice' that 'you must be able to mitigate the risk of identification until it is remote.'[25] The UK's ICO guidance on anonymisation remains relevant in context of research undertaken in the UK and thus subject to the UK Data Protection Act 1998. In considering the ADRN's robust process of de-identification, especially their use of Trusted Third Parties and other technical measures to prohibit re-identification, it would seem that even with such safeguards, the processing would *not* meet the higher standard of *irreversible* anonymisation proposed by the Article 29 Working Party in their anonymisation guidance. However when considering the Working Party's *2007* guidance on personal data, ADRN's arrangements seemingly *do* meet the requisite standard:

> In other areas of research ... re-identification of the data subject may have been excluded in the design of protocols and procedure, for instance because there is no therapeutical aspects involved ... [and] identification is not supposed or expected to take place under any circumstance, and appropriate technical measures (e.g. cryptographic, irreversible hashing) have been put in place to prevent that from happening. In this case, even if identification of certain data subjects may take place despite all those protocols and measures (due to unforeseeable circumstances such as accidental matching of qualities of the data subject that reveal his/her identity), the information processed by the original controller may not be considered to relate to identified or identifiable individuals taking account of all the means likely reasonably to be used by the controller or by any other person. Its processing may thus not be subject to the provisions of the Directive.[26]

While it is unclear how this 2007 guidance on personal data coincides with the Working Party's 2014 guidance on anonymisation (and thus the 'irreversible' standard referenced above), it would seem that if the ADRN ensures that re-identification is not to occur under any circumstances, with robust tech-

19 Blume (n 18) 34.

20 The term de-identification is used to refer to data that are anonymised on an individual level, where data may no longer be traced back to individual, nor directly or indirectly identifiable to the researcher. This is made possible through the use of Trusted Third Parties for the de-identification process, as supported by highly regulated access in safe settings, all being subject to the strictest of organisational measures, approvals and training. See: ADRN (n 8); ADRN (n 9).

21 ADRN (n 8).

22 DPD, recital 26.

23 Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (2014) WP216 29 5 <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>.

24 Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (n 23) 29.

25 The Information Commissioner's Office, 'Anonymisation: Managing Data Protection Risk Code of Practice' (2012) 6 <https://ico.org.uk/media/1061/anonymisation-code.pdf>.

26 Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data' (2007) 01248/07/EN WP 136 29 20 <http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf> (emphasis added).

DPD, Recital 26	GDPR, Commission, Recital 23	GDPR, Parliament, Recital 23	GDPR, Council, Recital 23
[Whereas] the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable. [DPD, recital 26 (emphasis added).]	The principles of data protection should not apply to data rendered anonymous in such a way that the data subject is no longer identifiable. (GDPR, Tripartite Version, Commission, recital 23.)	The principles of data protection should therefore not apply to anonymous data rendered anonymous in such a way that the data subject is no longer identifiable, which is information that does not relate to an identified or identifiable natural person.(GDPR, Tripartite Version, Parliament, recital 23.)	The principles of data protection should therefore not apply to anonymous information, that is information which does not relate to an identified or identifiable natural person or to data rendered anonymous in such a way that the data subject is not or no longer identifiable.(GDPR, Tripartite Version, Council, recital 23.)

Table 1

nical measures being put in place to prevent this from happening, the processing of such de-identified administrative data would not be subject to the DPD. (Irrespective of such processing being outwith the scope of the DPD, the initial transfer of data is still subject to *administrative* law i.e. public sector bodies' governing legislation. Even where data are de-identified prior to access being granted to a researcher, a public organisation must substantiate the legal grounds that support the transfer of data in the first place.[27])

2. A stricter standard of identifiability

Given that current access to and the use of administrative data in research is based upon the fact that such data are effectively de-identified (and thus outwith the scope of the DPD and thus the UK's Data Protection Act 1998), any changes made to the standard of anonymisation, from the current legal position, are of crucial concern. While anonymous data would remain outwith the scope of regulation under the GDPR, the standard, which determines when data are identifiable, and in turn, determines what are effectively non-identifiable, or anonymised data, is stricter than under the DPD. Recital 23, which stipulates the exclusion of anonymous data from the GDPR's scope, is almost identical to Recital 26 of the DPD in each of draft of the proposed legislation: see Table 1.

As under the DPD, the standard of anonymisation is directly related to the concept of identifiability, which determines what are or are not personal data. Under the DPD, data are considered identifiable by taking into account 'all the means likely reasonably to be used either by the controller or by any other person to identify the said person', a standard reflected in the Article 29 Working Party's guidance on anonymisation[28] and personal data[29] as well as that provided by the UK's ICO.[30] However, Recital 23 of the GDPR offers a more robust standard of identifiability (notably in Parliament and the Council's drafts): see Table 2.

Both Parliament and the Council's draft Recital 23 curtail current understandings of identifiability, and thus anonymisation, under both the DPD and UK data protection law. While the Commission's draft Recital 23 essentially transposes the current legal position on identifiability,[31] Parliament's draft extends this concept to include data which not only *identifies*

27 For consideration of the legal grounds which justify the use of administrative data for research under current law in the UK see: Laurie and Stevens (n 13). See also: Ministry of Justice, 'The Data Sharing Protocol: Annex H, Legal Guidance on Data Sharing' (27 July 2012) <http://webarchive.nationalarchives.gov.uk/20150730125042/http://www.justice.gov.uk/downloads/information-access-rights/data-sharing/annex-h-data-sharing.pdf>; The Law Commission, 'Data Sharing Between Public Bodies - A Scoping Report' (2014) <http://lawcommission.justice.gov.uk/docs/lc351_data-sharing.pdf>.

28 Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (n 23).

29 Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data' (n 26).

30 The Information Commissioner's Office (n 25) 6; The Information Commissioner's Office, 'Determining What Is Personal Data' (2012) 8 <https://ico.org.uk/media/for-organisations/documents/1554/determining-what-is-personal-data.pdf>.

31 Compare DPD, recital 26 to GDPR, Tripartite Version, Commission, recital 23.

DPD, Recital 26	GDPR, Commission, Recital 23	GDPR, Parliament, Recital 23	GDPR, Council, Recital 23
[Whereas], to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person. [DPD, recital 26 (emphasis added).]	To determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the individual. (GDPR, Tripartite Version, Commission, recital 23.)	To determine whether a person is identifiable, account should be taken of all the means likely reasonably likely to be used either by the controller or by any other person to identify or single out the individual directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the individual, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration both available technology at the time of the processing and technological development. (GDPR, Tripartite Version, Parliament, recital 23.)	To determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the individual directly or indirectly. To ascertain whether means are reasonable likely to be used to identify the individual, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration both available technology at the time of the processing and technological development. (GDPR, Tripartite Version, Council, recital 23.)

Table 2

a natural person but also which *singles them out*, directly or indirectly. While the Council's text differs, as it does not mention the words 'single out', their text specifically categorises *pseudonymous* data as identifiable to the extent that such data 'could be attributed to a natural person by the use of additional information'.[32] Pseudonymisation is characterised by the Article 29 Working Party on the basis that it still permits the 'singling out' of individuals, despite its use of de-identification techniques.[33] On this interpretation, the Council's version of Recital 23 may be considered a less explicit inclusion of 'singling out' which nevertheless would have the same impact on research carried out using de-identified, individual level records. A compromise position is likely to result from the Parliament's draft and Council's general position, which, as a result, is likely to enhance the standard of identifiability above current understandings that only revolve around whether data *identify* an individual, not single them out.

a. Singling out of individuals

The characterisation of data, which 'singles out' an individual, as personal data, represents a drastic enlargement of the concept of personal data and identifiability; something called for by privacy scholars[34] in response to ever evolving and sophisticated techniques in data linkages, the emergence of big data[35] and the acknowledged fallibility of anonymisation.[36]

32 GDPR, Tripartite Version, Council, recital 23.

33 Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (n 23) 10.

34 For example: Paul M Schwartz and Daniel J Solove, 'The PII Problem: Privacy and a New Concept of Personally Identifiable Information' (2011) 86 New York University Law Review 1814, 1877–1888.

35 Paul Ohm, 'The Underwhelming Benefits of Big Data' (2013) 161 University of Pennsylvania Law Review Online 339; Solon Barocas and Helen Nissenbaum, 'Big Data's End Run around Anonymity and Consent' in Julia Lane and others (eds), *Privacy, Big Data, and the Public Good: Frameworks for Engagement* (Kindle, Cambridge University Press 2014); Judith Rauhofer, 'Round and Round the Garden? Big Data, Small Government and the Balance of Power in the Information Age' (2014) 2014 University of Edinburgh, School of Law Research Paper Series <http://www.research.ed.ac.uk/portal/en/publications/round-and-round-the-garden(96582df3-858a-4e1f-9172-01124219c0c0).html>.

36 Arvind Narayanan and Vitaly Shmatikov, 'De-Anonymizing Social Networks', *30th IEEE Symposium on Security & Privacy* (2009) <https://www.cs.utexas.edu/~shmat/shmat_oak09.pdf>; Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2009) 57 UCLA Law Review 1701.

This approach to 'singling out' has been advocated by the Article 29 Working Party. After the general approach to the GDPR was agreed on 15 June 2015, the Working Party wrote to the European Commission urging them to retain the notion of 'singling out' in the final version of the GDPR as the explicit terminology was not retained from the Parliament's text:

> To ensure the general objective of maintaining a high-level of protection of personal data is upheld, personal data should be defined in a broad manner in line with technological evolution. The definition of personal data should therefore take into account the situation in which people can be "singled out" on the basis of identifiers.[37]

It remains unclear how the inclusion of 'singling out' in the concept of identifiability will impact upon research where access to data is strictly governed and re-identification is both prohibited and secured against through appropriate technical measures. Although ADRN accredited researchers will never have access to data, which has directly or indirectly identifiable information, it could be that because such data are at the *individual* level, rather than *aggregate* level,[38] that it may come within scope of the GDPR as data which 'singles out' an individual. This is so even if we consider the caveat carved out for research processing in Parliament and the Council's draft Recital 23 which stipulates that the GDPR does not concern the processing of *anonymous* information/data for statistical and research purposes.[39]

While there is no explicit connection made in the GDPR between 'anonymous data' on the one hand, and either aggregate level or individual level de-identified data, it is likely that in the inclusion of 'singling out' by Parliament and 'pseudonymous data' by the Council in Recital 23, that individual-level de-identified data may indeed come within the scope of the GDPR. This interpretation is supported by the fact that the Article 29 Working Party does connect individual level records that are de-identified to pseudonymisation and processing that 'singles out' individuals, which are considered to come within full scope of even the current law.[40]

The term 'singling out' can be traced at least back to the Article 29 Working Party's 2013 report on obtaining consent for cookies.[41] In pertinent part, they provide:

> Users should also be offered a real choice regarding tracking cookies. Such tracking cookies are generally used to follow individual behaviour across websites, create profiles based on that behaviour, infer interests, and take decisions affecting people individually. When tracking cookies are being used to single out people in this way, they are likely to be personal data.[42]

Concerns with causing individual impact on the basis of singling individuals out in regards to their behaviour, interests etc. is valid and should be regulated in context of the use of cookies. However, this concept of 'singling out' can and should be materially distinguished from the processing of de-identified data for research by ADRN accredited researchers. The *purpose* for using data is an important consideration (despite the Working Party's later contentions that 'it does not matter what the intentions are of the data controller or recipient'.[43]) In the case of research undertaken via the ADRN, it is *not* their purpose to track behaviour or individual interests in a way that affects individuals or will result in the taking of decisions affecting them – in fact such actions are prohibited. Such actions would be entirely inappropriate in context of the use of administrative data for research purposes.

It is not clear that Recital 23's inclusion of 'singling out' reflects concerns over individual impact, calling into question what its inclusion is meant to provide over and above provisions relating to profiling in e.g. Article 3(2)(b) and Article 20 of the GDPR. The Article 29 Working Party suggested in 2012 that 'singling

37 'Letter from the Article 29 Data Protection Working Party on Trilogue to Ms Vera Jourova, Commissioner for Justice, Consumers and Gender Equality of the European Commission' (17 June 2015) 2 <http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2015/20150617_letter_from_the_art29_wp_on_trilogue_to_msjourova.pdf>.

38 Anonymised records on the individual level versus aggregate, population level statistics.

39 GDPR, Tripartite Version, Parliament and Council, recital 23.

40 Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (n 23) 10.

41 Article 29 Data Protection Working Party, 'Working Document 02/2013 Providing Guidance on Obtaining Consent for Cookies' (2013) WP 208 <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp208_en.pdf>.

42 Article 29 Data Protection Working Party, 'Working Document 02/2013 Providing Guidance on Obtaining Consent for Cookies' (n 41) 6.

43 Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (n 23) 10.

out' be incorporated into Recital 23 of the Commission's original proposal for the GDPR as:

> One of the main conclusions of this analysis is that a natural person can be considered identifiable when, within a group of persons, he or she can be distinguished from other members of the group and consequently be *treated differently*. It is therefore suggested to clarify in Recital 23 and Article 4 that the notion of identifiability also includes singling out in this way.[44]

Again, the Working Party is concerned with singling out when it would result in *treating individuals differently*, when 'singling out' would result in impact upon individuals. However, this conception of singling out is disconnected from their 2014 opinion on anonymisation which instead focuses on the general risk of re-identification; in 2014, singling out is instead connected directly to pseudonymisation such that 'pseudonymised data cannot be equated to anonymised information as they continue to allow an individual data subject to be singled out and linkable across different data sets'.[45] They define 'singling out' as 'the possibility to isolate some or all records which identify an individual in the dataset'.[46] Importantly, no mention is made of individual impact and rather they emphasise that '"identification" not only means the possibility of retrieving a person's name and/or address, but also includes potential identifiability by singling out, linkability and inference.'[47] Their 2014 interpretation of 'singling out' is detached from the original (and appropriate) concerns posed regarding individual impact; the harm which the Article 29 Working Party would seek to prevent in regards to 'singling out' in their 2012 report on the GDPR and 2013 cookie consent report.

The Article 29 Working Party's more recent approach to 'singling out' was seemingly adopted into the wording of Parliament, and less explicitly in the Council's text for Recital 23, representing a departure from the more proportionate approach offered under the DPD and previously by the Working Party itself. Notably, in the Working Party's 2007 guidance on personal data, it was acknowledged that there would be circumstances, notably in the research context, where similar to the arrangements governing the ADRN, 'identification is not supposed or expected to take place under any circumstance, and appropriate technical measures ... have been put in place to prevent that from happening.'[48] Such circumstances were deemed to be *outwith the scope* of the DPD because the risks posed to individuals were effectively eliminated to the point of a less than remote possibility.[49] 'Singling out' as reflected in Recital 23 of the GDPR is disconnected from the original and appropriate concerns regarding singling out which impacts upon individuals in a significant way. The crucial effect of this disconnect is that the full force of the GDPR would apply to processing that would previously have been deemed 'safe' and outwith the scope of the GDPR, despite the implementation of appropriate technical and organisational measures to ensure individual impact and re-identification will *not* occur.

Recital 23 clearly demonstrates valid and acknowledged concerns with re-identification, in consideration of the well-known limits to anonymisation.[50] Nonetheless, the risks posed by the type of arrangements governing access and use of data by ADRN accredited researchers, in tightly constrained, secured (monitored settings), are *not* to be equated with the high profile cases of data breaches, where supposedly anonymous data were publicly released and later re-identified.[51] In fact, evidence suggests that use of data in research settings is remarkably safe in contrast to the more familiar and publicised instances of data breaches.[52] Further consider the UK's ESRC Annual Report 2013-2014 (which funds the type of re-

44 Article 29 Data Protection Working Party, 'Opinion 01/2012 on the Data Protection Reform Proposals' (2012) 5 <http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp191_en.pdf> (emphasis added).

45 Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (n 23) 10.

46 Ibid.

47 Ibid.

48 Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data' (n 26) 29.

49 Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data' (n 26).

50 Narayanan and Shmatikov (n 36); Ohm (n 36); Schwartz and Solove (n 32); Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (n 23).

51 Such as, Ryan Singel, 'Netflix Spilled Your Brokeback Mountain Secret, Lawsuit Claims' <http://www.wired.com/2009/12/netflix-privacy-lawsuit/>; Michael Barbaro and Tom Zeller, 'A Face Is Exposed for AOL Searcher No. 4417749 - New York Times' <http://www.nytimes.com/2006/08/09/technology/09aol.html?pagewanted=all&_r=0>.

52 For example see the findings of a recent Wellcome Trust report on data linkages in the health research sector: Public Health Research Data Forum, 'Enabling Data Linkage to Maximise the Value of Public Health Research Data' (Wellcome Trust 2015) 23, 29, 33, 39, 44, <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTP056860.htm>.

search discussed in this article) whereby *zero* 'protected personal data-related incidents' occurred which required formal reporting to the ICO during the period of 2011-2014.[53]

When considering the perspective of publicly funded and publicly beneficial social sciences research, concerns with 'open data' and unsecure use are not an issue and should not be conflated with the use of de-identified data, which are only ever accessed and used under strictly secured and monitored settings, subject to enforceable governance arrangements. It is entirely appropriate, and a welcome addition for the protection of individuals, that data processing which could be used in decisions significantly affecting them is now more strictly regulated in context with Article 3 and Article 20 of the GDPR. However, the approach taken to 'singling out' in Recital 23 does not allow for proportionate, risk-based consideration of differences between types of data controllers (and their purposes for processing), processing techniques, de-identification methods and other safeguards used to protect individuals' rights and interests in their data.

Consideration of context is vital to proportionate implementation of a regulation with direct effect across twenty-eight Member States, which will apply to a variety of necessarily different processing situations.[54] The disproportionate approach taken to singling out in Recital 23 may have the opposite effect in acting as a *disincentive* to use privacy protecting measures which require substantial investment of resources and time, including the use of Trusted Third Parties in rigorous de-identification techniques and the deployment of organisational measures such as safe settings for controlled, monitored access to data. It is unclear what incentives *would* remain, to implement costly safeguards, if such processing would be equated to far less secure processing of fully identifiable personal data, for less publicly beneficial reasons. Curiously, incentives do remain carved out for data controllers operating in the *private* sector given that the use of cookies and other online identifiers to create profiles on individuals (and thus single them out) would seemingly be *outwith* the scope of the GDPR so long as said cookies etc. are not combined with 'unique identifiers' or other information in order to identify an individual or make them identifiable.[55] It is unclear what reasoning could legitimately support such concessions, in contrast with the stricter treatment of secure and publicly beneficial processing for research where profiling of individuals in this way is made contrary to the requisite governance protocols. The lack of proportionality and risk-based assessment permitted by this approach, which impacts upon the most fundamental of data protection issues in determining what processing is or is not within the scope of the law, has the most potential to disrupt currently robust legal and technical arrangements for research.

b. Are pseudonymous data personal data?

While Recital 23 both explicitly (Parliament's draft[56]) and implicitly (the Council's draft[57]) calls into question the use of pseudonymous data through the concept of 'singling out', changes to Article 4 of the GDPR unambiguously creates a new subset of personal data in the formal addition of pseudonymous data.[58] Currently, the DPD defines personal data in reference to the definition of a data subject under Article 2(a):

> "[Personal] data" shall mean any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.[59]

Now contrast the current definition to those provided in the drafts to GDPR: see Table 3.

While the Commission's original proposal does not significantly expand upon the current legal posi-

53 ESRC, 'Economic and Social Research Council Annual Report and Accounts 2013-2014' (2014) 39 <http://www.esrc.ac.uk/_images/ESRC%20AR_tcm8-31173.pdf>.

54 Considering the impact of a regulation imposed on both private and public sector processing see: Blume (n 18).

55 While such processing would remain within the scope of the E-Privacy Directive art 5(3), from a data protection context, profiling on this basis would remain largely uninhibited by virtue of Recital 24, contrary to the concerns raised with regard to 'singling out' and its potential impact on individuals. GDPR, Tripartite Version, Council, recital 24.

56 GDPR, Tripartite Version, Parliament, recital 23.

57 Recital 23 of the Council's draft specifically provides that 'Data including pseudonymised data, which could be attributed to a natural person by the use of additional information, should be considered as information on an identifiable natural person.' See GDPR, Tripartite Version, Council, recital 23.

58 GDPR, Tripartite Version, Parliament, art 4(2)(a) and Council, art 4(3)(b).

59 DPD, art 2(a) (emphasis added).

GDPR, Commission, Article 4(1),(2)	GDPR, Parliament, Article 4(2)	GDPR, Council, Article 4(1)
(1) '[Data] subject' means an identified natural person or a natural person who can be identified, directly or indirectly, by means reasonably likely to be used by the controller or by any other natural or legal person, in particular by reference to an identification number, location data, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person; (2) 'personal data' means any information relating to a data subject; [GDPR, Tripartite Version, Commission, art 4(1),(2) (emphasis added).]	(2) '[Personal] data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable 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, unique identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social or gender identity of that person; [GDPR, Tripartite Version, Parliament, art 4(2) – Parliament deleted art 4(1) (emphasis added).]	(1) '[Personal] data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable 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, online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person;[GDPR, Tripartite Version, Council, art 4(1) (emphasis added).]

Table 3

tion (only adding location data, online identifiers and genetic data), both Parliament and the Council make noteworthy changes to expand the concept. First, both drafts remove the standard of 'by means reasonably likely to be used by the controller or by any other natural or legal person' an important threshold of proportionality for determining the identifiability of data (although retaining it in the guidance provided by Recital 23). Second, Parliament's draft Article 4(2) adds to the definition *indirect* identification by *unique identifiers*, which when read alongside Recital 23, could implicate de-identified data that are key coded and provided to researchers.

It is especially unclear how Parliament or the Council's definition of 'personal data' works alongside (i) Parliament's new category of 'pseudonymous data' in Article 4(2)(a), or (ii) Council's explicit inclusion of 'pseudonymous data' in Recital 23 and mention of pseudonymisation in Article 4(3)(b). Parliament provides an ambiguous definition for pseudonymous data without any reference to 'unique identifiers' or 'singling out' of individuals, both of which implicitly reference the process of pseudonymisation:

> "[Pseudonymous] data" means personal data that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution.[60]

In Recital 23 Parliament specifically defines identifiability on the basis of a data controller or any third party's ability 'to identify or *single out* the individual directly or indirectly'.[61] As discussed at length above, singling out individuals was indeed introduced in context of the process of *pseudonymisation* of data.[62] Thus it is unclear what Parliament's formal addition of pseudonymous data is intended to provide given that it stipulates such data *cannot* be attributed to a specific data subject without additional information – does this mean that such data, when 'kept separately and subject to technical and organisational measures to ensure non-attribution' are *outwith* the scope of the GDPR; or do pseudonymous data invoke an entirely different standard of protection? These questions certainly remain unresolved in consideration of the 'lesser' standard of protection seemingly offered to pseudonymous data in Parliament's draft as regards to e.g. the legitimate interests provisions (Recital 38), profiling (Recital 58(a) Article 10) etc.

Further inconsistencies arise when considering the Council's draft Article 4(3)(b) regarding pseudo-

60 GDPR, Tripartite Version, Parliament, art 4(2)(a).

61 GDPR, Tripartite Version, Parliament, recital 23 (emphasis added).

62 Article 29 Data Protection Working Party, 'Opinion 01/2012 on the Data Protection Reform Proposals' (n 44) 29; Article 29 Data Protection Working Party, 'Working Document 02/2013 Providing Guidance on Obtaining Consent for Cookies' (n 41) 29; Article 29 Data Protection Working Party, 'Opinion 05/2014 on Anonymisation Techniques' (n 23) 29.

nymisation when read alongside its provisions regarding pseudonymous data in Recital 23. The Council defines pseudonymisation as:

> [The] processing of personal data in such a way that the data can no longer be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organisational measures to ensure non-attribution to an identified or identifiable person.[63]

This may not create an explicit or formal category of pseudonymous data in the way Parliament did, but the effect is the same (and definition is essentially identical to that provided by Parliament). Therefore the Council's definition of 'pseudonymisation' also presents a pseudo-category of personal data, leaving it unclear what standard applies. The Council's Recital 23 provides that 'Data including pseudonymised data, which could be attributed to a natural person by the use of additional information, should be considered as information on an identifiable natural person.'[64] This characterises pseudonymous data *as* personal data; however, the Council perplexingly adds that pseudonymisation is a risk mitigating measure for protecting data subjects in Recital 23(a) and then goes on to promote the need to incentivise use of pseudonymisation in Recital 23(c):

> In order to create incentives for applying pseudonymisation when processing personal data, measures of pseudonymisation whilst allowing general analysis should be possible within the same controller when the controller has taken technical and organisational measures necessary to ensure that the provisions of this Regulation are implemented, taking into account the respective data processing and ensuring that additional information for attributing the personal data to a specific data subject is kept separately.[65]

Again the question remains: What incentives are there to pseudonymise if pseudonymous data, even those subject to the most robust technical security and organisational measures, come within the full scope of the GDPR? While Recital 23(c) may be partially reminiscent of the more proportionate approach taken to anonymisation and personal data in the DPD, UK's ICO and Article 29 Working Party's guidance on personal data, the effect is still that pseudonymous data would be treated – at least partially – like fully identifiable personal data. As stated above, outwith the research context there may indeed remain incentives to pseudonymise data, i.e. in context with a data controller's use of online identifiers such as cookies etc. given the concessions made in Recital 24. However, in the context of *research*, the changes proposed by Parliament or the Council in regards to personal data and pseudonymous data would require data controllers to demonstrate full compliance with the provisions of the GDPR, subject to any exceptions for research use and/or for pseudonymous data. This presents a disproportionate consideration of the likely risks at stake and sanctions an approach to regulation that lacks context sensitivity. Moreover, it displays a fundamental misunderstanding of the standard to which data are de-identified in the context of social sciences research, such as that supported by the ADRN.

3. Legitimising research use of administrative data

In the event that even robustly de-identified data for research are treated as personal data (or pseudonymous data, subject to varying standards of protection), data controllers involved in social sciences research must be able to legitimise their receipt and subsequent use of data under the GDPR. In a positive development for research, the GDPR introduces a legal ground that explicitly justifies the processing of personal data on the basis of research (or scientific) purposes: see Table 4.

Every use of personal data must be lawful. In order to be lawful, use of personal data must be justified upon one or more legal grounds provided in data protection law. In the research context this typically involves obtaining consent from data subjects or demonstrating that the processing was necessary for the legitimate interests of the data controller, so long as the use does not prejudice data subjects' freedoms, rights or interests.[66] Currently, the use of personal data cannot be legally justified on the basis of conducting research on its own. That is, use of personal

63 GDPR, Tripartite Version, Council, art 4(3)(b).

64 GDPR, Tripartite Version, Council, recital 23(a).

65 GDPR, Tripartite Version, Council, recital 23(c).

66 DPA, Schedule 2, para 1, 6.

GDPR, Commission, Article 6(2)	GDPR, Parliament, Article 6(2)	GDPR, Council, Article 6(2)
Processing of personal data which is necessary for the purposes of historical, statistical or scientific research shall be lawful subject to the conditions and safeguards referred to in Article 83.[GDPR, Tripartite Version, Commission, art 6(2).]	Processing of personal data which is necessary for the purposes of historical, statistical or scientific research shall be lawful subject also to the conditions and safeguards referred to in Article 83.[GDPR, Tripartite Version, Parliament, art 6(2).]	Processing of personal data which is necessary for archiving purposes in the public interest, or for historical, statistical or scientific purposes shall be lawful subject also to the conditions and safeguards referred to in Article 83.[GDPR, Tripartite Version, Council, art 6(2).]

Table 4

data for research must be legally justified on the basis of consent, or another legal ground as provided under Article 7 of the DPD.

Under the Commission and Parliament's draft Article 6, 'scientific research' becomes a legal ground for processing personal data, whereas under the Council's Article 6, scientific purposes are instead referred to.[67] Furthermore, legal grounds are provided under Article 9 to legitimise processing of special categories of personal data on the basis of research/ scientific purposes.[68] However, legal justification remains conditioned upon the basis of fulfilling the provisions of Article 83, which specifically considers the use of data for historical/archival, statistical and research/scientific purposes. Article 83 is to be read in tandem with Article 6(2). Under the Commission and Parliament's drafts, Article 83 is worded such that personal data may only be processed for research under specific conditions, whereby the Council's draft is more permissive and instead provides that Article 83 applies *where* personal data are processed for research, etc.[69]

The Commission's and Parliament's drafts provide that data controllers may only rely upon research as a legal justification for processing personal data if:

> (a) These purposes cannot be otherwise fulfilled by processing data which does not permit or not any longer permit the identification of the data subject;
>
> (b) Data enabling the attribution of information to an identified or identifiable data subject is kept separately from the other information *under the highest technical standards, and all necessary measures are taken to prevent unwarranted re-identification of the data subjects.*[70] (Emphasis showing the amendments made by Parliament.)

The Commission's version of Article 83 can be distinguished from above as it only requires separation of identifying and non-identifying data to the extent that research can be fulfilled in this manner[71] and makes specific provision for the disclosure of personal data where necessary to present research findings (important considering many UK funding bodies' open access policies).[72] This more proportionate approach is not replicated in Parliament's draft, which does not provide for publication of research results and insists that identifiable information always remain separate from non-identifiable data.[73] The health research community has been particularly vocal in their concerns over Parliament's draft Article 83, as 'This amendment makes the exemption from consent for the use of health data in research very narrow, which will prevent valuable research that is currently legal.'[74] As applied to social sciences research, different concerns are raised, and the continued separation of identifiable data from de-identified research data is already standardised in the proce-

67 GDPR, Tripartite Version, Commission and Council, art 6(2).

68 GDPR, Tripartite Version, art (9)(i).

69 GDPR, Tripartite Version, art 83.

70 GDPR, Tripartite Version, Commission and Parliament, art 83(1)(a),(b) (emphasis added to show Parliament's amendment).

71 GDPR, Tripartite Version, Commission, art 83(1)(b).

72 Publication of research results implicating personal data would be allowed under the Commission's proposed Article 83(2(b) so long as the rights and interests of the individuals do not override this. GDPR, Tripartite Version, Commission, art 83(2)(b).

73 GDPR, Tripartite Version, Parliament, art 83(1)(b).

74 Dr Beth Thompson, Policy Adviser, Wellcome Trust, 'Protecting Health and Scientific Research in the Data Protection Regulation: Position of Non-Commercial Research Organisations and Academics - December 2014' 19 <http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/WTP055584.pdf>.

dure for de-identification undertaken via the ADRN[75] (and likely many other research bodies).

Therefore, even under the more stringent requirements of Parliament, Article 83 is not insurmountable for social sciences research. The provisions can be interpreted as simply requiring that the use of *personal* data for research is strictly necessary – if the research can otherwise be conducted with de-identified data it should do so. Indeed, this standard of 'necessity' is not novel and is already required in several provisions of current data protection law e.g. DPD, Article 7(b)-(f).[76] Moreover, where de-identified data are used, Article 83(1)(b) merely signposts to the important role of the Trusted Third Party in social sciences research to maintain clear separation of identifiable information and to ensure adherence to robust security protocol for the transfer and subsequent use of de-identified data for research.[77]

Overall, the explicit recognition of research as a legitimate, legal ground for processing is a positive development for social sciences research (and other forms of research). However, it is important to note that a data controller's reliance on Article 6(2) is affected if the processing involves the *reuse* of data as opposed to data specifically collected for research purposes. Much research, including that of the ADRN, relies upon the reuse of *existing* data sets, which are originally collected for purposes *other than* research. Therefore, data controllers would need to demonstrate that their further processing for research is *compatible* with the original purposes for collection according to the principle of purpose limitation. Crucially, the Council's Article 5(1)(b) provides that 'further processing of personal data for archiving purposes in the public interest or scientific, statistical or historical purposes shall in accordance with Article 83 not be considered incompatible with the initial purposes'.[78] According to this approach, so long as a data controller complies with the requirements of Article 83 (discussed above) their reuse of data for research would automatically be considered 'compatible' and in compliance with the principle of purpose limitation. However it is crucial to consider whether the law is enough to justify compatibility in *all* cases: Should research (of *any* type, presumably also including commercial/market research) *always* be considered compatible with the purposes of original collection? Article 83 does provide certain safeguards to individuals in regards to research, making the determination of compatibility contingent upon these provisions being satisfied. But as made abundantly clear from recent experiences in the health research context, namely care.data in the UK,[79] legal sanction does *not* equate to social licence and public acceptability. More careful consideration is needed to affect a proportionate balance between the public interest in *both* the protection of individual privacy and in certain research uses of data.

Nevertheless, data controllers can rely upon Article 6(2) as their lawful justification for reuse of data for research, subject to a determination of compatibility and compliance with Article 83. In cases where data controllers are *unable* to comply with the requirements of Article 83, another lawful basis under Article 6 would need to be satisfied for the processing to be considered lawful when data are reused for research. In such cases, obtaining consent may become important and the GDPR does make it increasingly difficult to obtain *valid* consent, especially in the research context, given the specificity required and the impact of rights to withdrawal.[80] For consent to be considered valid under the Commission and Parliament's draft it must be 'freely given specific, informed and explicit'[81] while the Council provides that consent must be 'freely-given, specific, and informed'.[82] Moreover, the Commission and Parliament provide in Recital 25 that 'Consent should be given explicitly by any appropriate method enabling a freely given specific and informed indication of the data subject's wishes'[83] in contrast to 'un-

75 ADRN (n 8).

76 For example consider Article 7 of the current Data Protection Directive 95/46/EC – all legal grounds for processing, with the exception of obtaining individual consent, are conditioned upon the processing being *necessary* for the particular purposes relied upon. This approach is mirrored in the UK's DPA 1998, Schedule 2.

77 Although, it remains unclear what is meant by 'the highest technical standards.' GDPR, Tripartite Version, Parliament, art 83(1)(b).

78 GDPR, Tripartite Version, art 5(1)(b).

79 Olivia Solon, 'The Communication of Care.data to Patients Has Been an Absolute Shambles' <http://www.wired.co.uk/news/archive/2014-02/07/care-data-terrible-communication>; Pam Carter, Graeme T Laurie and Mary Dixon-Woods, 'The Social Licence for Research: Why Care.data Ran into Trouble' [2015] Journal of Medical Ethics <http://jme.bmj.com/content/early/2015/01/23/medethics-2014-102374.abstract>.

80 GDPR, Tripartite Version, art 7.

81 GDPR, Tripartite Version, Commission and Parliament, art 4(8).

82 GDPR, Tripartite Version, Council, art 4(8).

83 GDPR, Tripartite Version, Commission and Parliament, recital 25.

ambiguous' consent provided by the Council.[84] All drafts require the data controller to demonstrate that consent was obtained under these more rigorous terms[85] and clearly distinguish, and in practice separate, requests for consent for different purposes.[86] Parliament's draft Article 7 diverges in that consent would be considered void if obtained without clear separation of consent for different purposes, which would threaten any prospect for broad consent.[87]

A more positive outcome for research is reflected in the Council's Recital 25 which allows broad consent to scientific research areas:

> It is often not possible to fully identify the purpose of data processing for scientific purposes at the time of data collection. Therefore data subjects can give their consent to certain areas of scientific research when in keeping with recognised ethical standards for scientific research. Data subjects should have the opportunity to give their consent only to certain areas of research or parts of research projects to the extent allowed by the intended purpose and provided that this does not involve disproportionate efforts in view of the protective purpose.[88]

This official recognition of the value of broad consent is crucial for longitudinal studies, and beneficial to social sciences research overall. The methods used to produce scientifically sound and ethically robust social sciences research are not as formulaic or defined at the outset (as would be the case in biomedical or clinical research).[89] These factors make informed consent difficult to obtain, especially under the enhanced standards provided in Article 7 of the GDPR. There are fundamental differences between clinical and biomedical research and social sciences research; whereas the former may involve serious interventions (physical) on the individual, and therefore the risks posed are often more severe,[90] research which instead solely relies upon the reuse of data originally collected elsewhere (which may also include health research) poses substantially different types of risk (and often less risk) to individuals, especially when considering the robust de-identification processes and other safeguards implemented by organisations like the ADRN. Therefore the introduction of a specific legal ground to justify the processing of data on the basis of research itself, in combination with the declaration of compatibility in Article 5(1)(b), is especially beneficial to social sciences research that relies upon the reuse of de-identified administrative data.

III. Concluding thoughts

The implications of the GDPR are not wholly negative for social sciences research. In fact, there is increased and explicit recognition of the importance of research (helpfully defined to include 'fundamental research, applied research, and privately funded research'[91]) throughout the GDPR. This is a positive development for social sciences research for a number of reasons. Unlike under the DPD, the GDPR explicitly recognises research as a valuable form of processing in itself, demonstrated through the addition of a legal ground on the basis of research and in several new provisions dedicated to research e.g. Recital 25(aa), Article 83 etc. Furthermore, the addition of a separate legal ground for research diminishes the prominence and 'fetishisation of consent', which has been considered an impoverished way of protecting individual's rights and other interests in their data, as consent is not necessarily indicative of data being adequately protected nor data subjects' rights and interests being respected.[92] In fact:

> [It] is conceivable that processing can take place lawfully (with a valid legal basis) but without what

84 GDPR, Tripartite Version, Council, recital 25 and art 6(1)(a).

85 GDPR, Tripartite Version, art 7(1).

86 GDPR, Tripartite Version, art 7(2).

87 GDPR, Tripartite Version, Parliament, art 7(2).

88 GDPR, Tripartite Version, Council, recital 25(aa).

89 David Erdos, 'Constructing the Labyrinth: The Impact of Data Protection on the Development of "Ethical" Regulation in Social Science' (2012) 15 Information Communication and Society 104; David Erdos, 'Systematically Handicapped? Social Research in the Data Protection Framework' (2011) 20 Information Communication and Society 83; David Erdos, 'Stuck in the Thicket? Social Research under the First Data Protection Principle' (2011) 19 International Journal of Law and Information Technology 133; Robert Dingwall, 'The Ethical Case against Ethical Regulation in Humanities and Social Science Research' (2008) 3 21st Century Society: Journal of the Academy of Social Sciences 1.

90 Dingwall (n 89).

91 GDPR, Tripartite Version, recital 126.

92 Graeme Laurie and Emily Postan, 'Rhetoric or Reality: What Is the Legal Status of the Consent Form in Health-Related Research?' (2012) Medical Law Review <http://medlaw.oxfordjournals.org/content/early/2012/10/09/medlaw.fws031.abstract>; Graeme Laurie and Shawn Harmon, 'Through the Thicket and Across the Divide: Successfully Navigating the Regulatory Landscape in Life Sciences Research' (2013) 30 University of Edinburgh, Research Paper Series, SSRN <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2302568>.

> the authors would consider 'appropriate' protections. Thus, in the case of data subject consent as a legitimate ground for processing personal data, the data subject's self-determined decision to agree to processing his own data may prevail over a possible lack of protection for his personal data.[93]

Dispelling the infallibility of consent in research governance is not only beneficial from the standpoint of effectuating more meaningful and dynamic protection of individuals' interests and rights. It is also of particular benefit to the social sciences research community given the often criticised, but legally unchallenged, imposition of a biomedical and clinical style of research governance which favours obtaining informed consent.[94] Unlike the prescribed nature of much biomedical and clinical research, social sciences research does not presuppose the outcomes of a particular project; rather, outcomes evolve and emerge throughout.[95] This makes it particularly difficult to convey the level of information necessary for consent to be valid under current legal requirements and certainly so under the drafts being considered for the GDPR. Where consent would be appropriate to obtain for social sciences research, the general approach recently adopted by the Council navigates the issues in a more nuanced and proportionate manner.[96]

What remains more problematic for social sciences research is the enlargement of the concept of identifiability and thus personal data, which would treat pseudonymous data as a subset of personal data. Unlike the more proportionate approach provided for under the DPD and advocated earlier by the Article 29 Working Party in its 2007 guidance on personal data, the GDPR would apply fully to the robustly de-identified data used by ADRN for research, even where 'identification is not supposed or expected to take place under any circumstance, and appropriate technical measures (e.g. cryptographic, irreversible hashing) have been put in place to prevent that from happening.'[97] This disrupts currently sanctioned understandings of anonymisation and identifiability in the UK, but also if one considers the Article 29 Working Party's own guidance on personal data.[98] While the singling out of individuals for the purpose of taking decisions or otherwise significantly them should indeed be regulated, the standard for identifiability offered in the GDPR is seemingly disconnected with this concern. It therefore remains unclear how the inclusion of 'singling out' will work alongside Parliament's ambiguous definition of pseudonymous data,[99] or indeed the inconsistencies posed by the Council's general approach to pseudonymisation which seemingly recognises the importance of incentivising privacy protecting mechanisms while still applying the full force of the GDPR upon such processing.[100]

It is welcome to see that the GDPR will formally recognise limits to anonymisation and the specific risks presented by pseudonymised, individual level data (as opposed to aggregated, population level data). However, the approach taken to pseudonymous data in the Parliament's and the Council's drafts lacks nuance and proportionate consideration of the necessarily varied circumstances under which data are de-identified, by different types of processes, for different purposes and by different data controllers. Whereas the current legal position technically allows for such context sensitive determinations, the GDPR would not. A more proportionate approach would allow consideration of the specific methods and procedures used by data controllers in ensuring the remote and unlikely chance of re-identification when individual level but robustly de-identified data are

93 Paolo Balboni and others, 'Legitimate Interest of the Data Controller New Data Protection Paradigm: Legitimacy Grounded on Appropriate Protection' (2013) 3 International Data Privacy Law 244, 246.

94 Rose Wiles and others, 'Informed Consent in Social Research: A Literature Review' (2005) NCRM 001 ESRC National Centre for Research Methods <http://eprints.ncrm.ac.uk/85/1/MethodsReviewPaperNCRM-001.pdf>; Rose Wiles and others, 'Informed Consent and the Research Process: Following Rules or Striking Balances?' (2007) 12 Sociological Research Online <http://www.socresonline.org.uk/12/2/wiles.html>; Rose Wiles and others, 'The Management of Confidentiality and Anonymity in Social Research' (2008) 11 International Journal of Social Research Methodology 417; Dingwall (n 89); Sarah Dyer and David Demeritt, 'Un-Ethical Review? Why It Is Wrong to Apply the Medical Model of Research Governance to Human Geography' (2009) 33 Progress in Human Geography 46; Erdos, 'Stuck in the Thicket? Social Research under the First Data Protection Principle' (n 89); Kristian Pollock, 'Procedure versus Process: Ethical Paradigms and the Conduct of Qualitative Research' (2012) 13 BMC Medical Ethics 1.

95 Erdos, 'Stuck in the Thicket? Social Research under the First Data Protection Principle' (n 89) 134-135.

96 GDPR, Tripartite Version, Council, recital 25.

97 Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data' (n 26) 29.

98 Article 29 Data Protection Working Party, 'Opinion 4/2007 on the Concept of Personal Data' (n 26) 20; The Information Commissioner's Office (n 25) 6; The Information Commissioner's Office (n 34) 27.

99 GDPR, Tripartite Version, Parliament, art 2(a).

100 GDPR, Tripartite Version, Council, recital 23 and art 4(3)(b).

used. Given that research is recognised as a valuable form of processing in itself, and by default 'compatible' with the original purposes of collection per Article 5(1)(b) of the Council's general approach, it is surprising that clearer distinctions are *not* made between pseudonymous data used for research versus other, more potentially risky purposes that are not likely to yield public benefit. The case for social sciences research remains in the balance as Europe awaits the outcome of final negotiations to the GDPR.