SUBMISSION TO THE INFORMATION COMMISSIONER'S OFFICE

on the subject of the draft Data Subject Access Rights Guidance

12 February 2020

Contents

Introduction

1. Open Rights Group (ORG) is a UK-based digital campaigning organisation working to protect fundamental rights to privacy and free speech online. With over 3,000 active supporters, we are a grassroots organisation with local groups across the UK.

2. We have responded to government and regulatory consultations in the area of data policy and data rights work for many years. Recent highlights include work on data protection enforcement, including challenges to online behavioural advertising practices, currently being investigated by the Information Commissioner. We have raised concerns about the lack of privacy protections relating to online age verification.

3. ORG is also a beneficiary of the ICO's grants programme, producing the *Data Rights Finder* tool and multiple reports on the state of data rights.

4. **We broadly support the proposed contents of ICO's subject access rights guidance, however believe the Guidance lacks some specific discussion and clarity, particularly concerning the current digital economy.** This response provides a range of areas in which we believe the Guidance can be usefully and easily improved.

Increase Relevance to the Digital Economy

5. The guidance clearly builds on the existing *Subject Access Code of Practice*, and many aspects of that document are commendable and important to restate going forwards. However, the Code of Practice and this proposed Guidance both share a weakness of being focussed too heavily on examples relating to 'traditional' data controllers (e.g. employers, hospitals), instead of adopting an approach with acknowledges the importance of new actors in the data economy, and the specific way access requests may differ in relation to them.

6. The following section breaks down some of these specific weaknesses, and suggests actionable changes to the Guidance to remedy them.

Inferences

7. The fact that opinions and inferences can qualify as personal data has been confirmed by the CJEU, which noted that the term 'any information' in the definition of personal data includes information that is 'not only objective but also subjective, in the form of opinions and assessments, provided that it 'relates' to the data subject'.[1] The test of whether data 'relates' to an individual is satisfied where it is linked to a person 'by reason of its content, purpose or effect'.[2]

8. Opinions and inferences formed of the data subject by the controller fall within the right of access. These inferences can range from quantitative or 'predictive' assessment of employment performance using manual or automated surveillance tools[3] to profiling of data subjects by information society services.[4] Access to these opinions and inferences is key to a variety of other rights and obligations in the data protection regime, such as rectification, objection, erasure, as well as the broad assessment of fairness and non-discrimination.[5] **Access rights are important pre-**

---

[1] Case C-434/16 *Peter Nowak v Data Protection Commissioner* ECLI:EU:C:2017:994 [34].
[2] ibid [35].
[3] See generally Ifeoma Ajunwa and others, 'Limitless Worker Surveillance' (2017) 105 Calif L Rev 735; Lilian Edwards and others, 'Employee Surveillance: The Road to Surveillance is Paved with Good Intentions' (SSRN Scholarly Paper, 18 August 2018).
[4] See generally Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For' (2017) 16 Duke Law & Technology Review 18.
[5] GDPR, recital 71.

requisites to checking legality, and providing them is key to effective oversight and the principle of transparency.[6]

9. While opinions of data subjects by third parties who are not the data controller should be treated, as Guidance describes, as personal data that relates to more than one person, opinions of the data subject derived through software *by the data controller* will not generally be subject to such considerations.

10. The limitations discussed in the literature and by the ICO on meaningful information about the logic of processing related to automated decision-making are irrelevant to the question of whether inferences fall within the right of access.

11. In practice, many controllers currently make inferences about individuals using arcane methods such as data embeddings, where an individual's data is transformed into a more abstract representation of hundreds of numbers representing their 'location' in a data space in relation to other data subjects. **In principle, even seemingly abstract inferred data like this should be included in access rights.** The Guidance already recognises that inferred data is subject to the right of access, but it should go further and clarify that **even if these inferences are complicated, or that the data controller cannot anticipate any use of them by the data subject, that does not exempt them from the access right.**

12. The guidance should be clear that such inferred data falls within Article 15(1)(g) of the GDPR: 'where the personal data are not collected from the data subject, any available information as to their source' [must be provided]. **This obliges data controllers to describe (and provide, where it is available) the data from which an inference was generated.**

Recommendation: Clarify that it is not for the data controller to judge the utility of any inferences (or other data) to the data subject, but it is an obligation to be transparent about their derivation, meaning and use.

Recommendation: In accordance with Article 15(1)(g), the data controller must specify the source of any personal data used to generate inferences.

Information to Aid Understanding of Data

13. The Guidance states, on page 33, that

> When providing a copy of the personal data requested, you are expected to give the individual additional information to aid understanding if the data is not in a form that they can easily understand. However, this is not meant to be onerous, and you are

[6] Jef Ausloos and Pierre Dewitte, 'Shattering One-Way Mirrors – Data Subject Access Rights in Practice' (2018) 8 International Data Privacy Law 4.

not expected to translate information or decipher unintelligible written notes.

14. We agree that it is disproportionate to, for example, decipher unintelligible written notes, **however it is important to state in this guidance what is clearly and unambiguously proportionate.**

15. It is common, in access requests, to receive

    (1) Variables whose names are encoded or have no clear and obvious meaning to a data subject.

    (2) Variables whose contents are coded with a code-book, for example numerically or with words. A simple example is the system of NRS social grades (A, B, C1, C2, D, E) commonly used in demographic statistics in the United Kingdom. However, many examples will be based on internally maintained definitions.

    (3) Numeric variables which are on a scale, where that scale is not provided.

16. **Data controllers must already be internally documenting what this data means under data protection law.** This can be implied from the principles of accuracy and accountability. Furthermore, to be collecting data where the organisation does not understand the purpose or meaning of it would be counter to the principle of data minimisation. **Consequently, information on variable purposes and meanings, codebooks and scales should already be available to the data controller.**

17. **The ICO should clarify that codebooks, variable descriptions and meanings and scales and contextual information will, as a result, not be considered onerous to provide, and indeed are required.**

Recommendation: Clarify that code-books, scales and variable meanings must be maintained inside a data controller, and therefore will not be onerous or disproportionate to provide to the data subject and will be required under an Article 15 request.

Bulk Requests

18. We strongly support the 'How should we deal with bulk requests?' recommendations in the Guidance and recommend they should not be weakened as a response to consultation responses.

Recommendation: Retain the Guidance around bulk requests.

Deleted Information

19. In many cases where an individual has 'pressed delete', the information remains stored in the system (such as deleted posts, or half-typed messages).[7]The section on deleted information should emphasise that this remains subject to the right of access.

Recommendation: Clarify that information the user believes they have 'deleted' but the service has retained remains in scope of the right of access.

Information about Recipients and Sources

20. **The Commissioner should elaborate on the requirement to disclose recipients of personal data**. Article 15(1)(a–f) encompasses the obligation on data controllers to provide additional information regarding the processing of data. Particularly important in relation to the data subject's ability to monitor the controller's compliance with data protection legislation as well as their ability to effectively exercise her other data subject rights are the right to know the recipients as well as the sources of the data undergoing processing. In line with these goals and building on the earlier position taken by A29WP and endorsed by the EDPB,[8] the provided information should include

    (1) the specific named sources of personal data

    (2) by default, the named recipients of the data subject's personal data. If 'categories' are to be provided, these must be specific and include the sector and location.

    Currently only a very small proportion of data controllers provides such data when requested.[9]

---

[7] Drew Harwell, 'Start a Post, Then Delete It? Many Websites Save It Anyway.', *Washington Post* (18 December 2018) <https://www.washingtonpost.com/technology/2018/12/18/start-post-then-delete-it-many-websites-save-it-anyway/> accessed 17 November 2019; Tony Romm, 'Facebook Says a New Bug Allowed Apps to Access Private Photos of up to 6.8 Million Users', *Washington Post* (14 December 2018) <https://www.washingtonpost.com/technology/2018/12/14/facebook-says-new-bug-allowed-apps-access-private-photos-up-million-users/> accessed 17 November 2019; Steven Englehardt and others, 'No Boundaries: Exfiltration of Personal Data by Session-Replay Scripts' (*Freedom to Tinker, Centre for Information Technology Policy, Princeton University*, 15 November 2017) <https://freedom-to-tinker.com/2017/11/15/no-boundaries-exfiltration-of-personal-data-by-session-replay-scripts/> accessed 17 November 2019.

[8] Article 29 Working Party, 'Guidelines on Transparency under Regulation 2016/679' (11 April 2018) 37.

[9] René LP Mahieu and others, 'Collectively Exercising the Right of Access: Individual Effort, Societal Effect' (2018) 7 Internet Policy Review.

21. A reading of Articles 13-15 makes it clear that as part of this information, the controller must also provide

    (1) the identities of all joint controllers;

    (2) the identities of all data processors;

    in particular as both will be clearly and explicitly known by the data controller with no additional effort required on their part.

Recommendation: Remind data controllers that they must name the sources from which data are received.

Recommendation: Specify that if a controller opts to use 'categories' rather than named recipients, these categories must be suitably granular, and include at least the sector and location of recipients.

Recommendation: Specify that joint controllers must be named regardless of whether they are 'recipients' of the data (on the basis of a restatement of Articles 13(1)(a) and 14(1)(a).

Recommendation: Specify that processors must be named, as they will be clearly know recipients for which there is no reason under the fairness and transparency principle to describe through categories.

Tailored Information Concerning Purposes and Lawful Bases

22. When a copy of data is received, the Guidance should specify that the data controller must indicate, for each category of data

    (1) The purposes for which this data is processed;

    (2) The lawful bases for processing, *broken down and specified by purpose*;

    (3) Where consent is used, information on when and how this consent was received;

    (4) Where legitimate interests are used, the description of the legitimate interests used;

    (5) Where legal obligation is used, the specific statutory or common law basis for such a legal obligation.

23. The data controller must not simply link to their privacy policy for points 1, 2, 4 and 5 of the above if it is not *manifestly clear* from such a policy which received data falls in which category, and which categories are covered by which lawful basis and purpose.

24. The current generation of 'Download My Data' tools employed by information society services and similar controllers do not contain information on lawful bases and purposes. The ICO should address this point, as many data subjects are redirected to these tools which only provide a small portion of the information they have the right to under Article 15, and *none* of the metadata.

Recommendation: Specify that in providing a copy of personal data, the controller must make it clear which parts of the data returned fall into which categories, for which purposes each category is being processed, and for which lawful base(s) each purpose is based. Clarify that providing a list that is not broken down by category (e.g. a copy-pasted version of Article 6 of the GDPR) is insufficient.

Recommendation: Specify that where consent is relied on, the time and manner of such consent must be provided to the data subject. Where legal obligation is used, the statutory or common law basis for this must be provided. Where legitimate interest is used, the legitimate interest itself must be provided.

Recommendation: Specify that while useful, 'download my data' tools must also provide metadata or supplementary information such as processing purposes or lawful bases.

Joint Controllership

25. The guidance states (p.5):

> Controllers are responsible for ensuring that SARs are complied with. If you are a joint controller, you need to have a transparent arrangement in place with your fellow joint controller(s) which sets out how you deal with SARs.

26. This is true, but omits to acknowledge that an individual can make a SAR against any joint controller in a joint controllership arrangement, and that following a string of recent, settled case-law, controllers, particularly online, may *de facto* be in much more joint controllership arrangements than they acknowledge. The CJEU has noted on multiple occasions that some joint controllers may never see a copy of the personal data being processed, as they only exert control of the means and ends.[10] This does not absolve them from having to acknowledge and respond to SARs.

27. This is particularly the case in, for example, the situation where a webpage has embedded a tracker from a tracking company. In *Fashion ID*, the Court held that a

[10] Case C-49/17 *Fashion ID GmbH & CoKG v Verbraucherzentrale NRW eV* ECLI:EU:C:2019:629 [82]; Case C-210/16 *Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein v Wirtschaftsakademie Schleswig-Holstein GmbH* ECLI:EU:C:2018:388 [38]; Case C-25/17 *Jehovan todistajat* [2018] ECLI:EU:C:2018:551 [69].

website was a joint controller in relation to trackers it embedded.[11] A similar argument has been made by the Court for operators of Facebook 'Fan Pages'.[12] Parallel arguments can easily be made for those embedding trackers in software such as apps.

28. Research, including that from ICO staff, has highlighted the large difficulties in identifying trackers present in apps and on websites.[13] **The Guidance should emphasise that data controllers making the conscious choice to embed third party trackers must have processes in place to properly fulfil access requests directed at them in relation to the data that these trackers collect.**

29. Because a website or app may be in tens or hundreds of distinct joint controllership arrangements in relation to a single data subject, it is **not in line with the principle of fairness, nor the concept of joint controllership, to ask a data subject to manually direct their request to these hundreds separately**. If an access request is directed to one joint controller, it is their responsibility to deliver it to the controller designated to undertake the task within their controllership arrangement.

Recommendation: Specify that a joint controllership arrangement does not permit a data subject to only be permitted to submit a SAR to certain joint controllers, giving the example of an embedded web tracker as per *Fashion ID*.

Recommendation: Specify that regardless of the responsibilities distributed within joint controllership arrangement, a SAR aimed at any joint controller must be forwarded to other joint controllers — the data subject cannot be asked to redirect their request.

Recommendation: Specify that the scope of a SAR relates to all data processed by the controller to whom the request is made in both their individual controllership capacity and any joint controllership capacities

Proposed Example: "*A news website uses an installed third-party tracker which gathers data about website visits against persistent identifiers. A data subject contacts the news website to ask for access to data collected by these trackers. As the website is a joint controller with the organisations who maintain the code for the trackers, it is the websites responsibility to pass the access request on to every tracking organisation they have a joint controllership arrangement with.*"

[11] *Fashion ID* (n 10).

[12] *Wirtschaftsakademie* (n 10).

[13] Reuben Binns and others, 'Third Party Tracking in the Mobile Ecosystem' in *Proceedings of the 10th ACM Conference on Web Science* (WebSci '18, New York, NY, USA, ACM 2018); Arjaldo Karaj and others, 'WhoTracks.Me: Shedding Light on the Opaque World of Online Tracking' [2018] arXiv:180408959.

Complexity Extensions for Information Society Services & Similar Controllers

30. Many data controllers in the digital economy routinely process a large amount of data about many individuals. The Guidance states

> What may be complex for one controller may not be for another – the size and resources of an organisation are likely to be relevant factors.

31. **Many data controllers are highly adept in the automated processing of personal data, and do so at a large scale. It is not in adherence to the principle of fairness to permit them to process data rapidly and swiftly for commercial purposes but to claim it is too complex to replicate the practice for the purposes of data protection.** Controllers with significantly complex day-to-day processing operations should as a general rule not be permitted to benefit from the complexity extension. This is mirrored in guidance endorsed by the EDPB in relation to data portability, that for the cases of information society services which specialise in automated data processing, 'there should be very few cases where the data controller would be able to justify a refusal to deliver the requested information, even regarding multiple data portability requests.'[14]

Recommendation: Clarify that for information society services which specialise in automated data processing, there should be few cases where the data controller would be able to claim a complexity extension.

Telemetry Data and Data Processing by Software Vendors

32. Agile software development, where software is delivered 'as-a-service' and is seen as in 'perpetual beta', has greatly increased the amount of personal data collected through the routine use of computing. **Telemetry data** is commonly part of this, effectively amounting to the detailed monitoring of exactly how a user uses a device or piece of software.

33. The risk associated with telemetry data has recently been recognised by the EDPS (in relation to Regulation (EU) 2018/1725) and Dutch Ministry of Justice and Security (in relation to the GDPR) in the context of Microsoft Office, who highlighted that this increased telemetry data collection was often done without appropriate lawful basis or DPIAs.[15]

---

[14] Article 29 Working Party, 'Guidelines on the Right to Data Portability (WP 242)' (13 December 2016) 9–10.

[15] European Data Protection Supervisor, 'EDPS Investigation into IT Contracts: Stronger Cooperation to Better Protect Rights of All Individuals', *EDPS/2019/07* (Brussels, 21 October 2019); Letter from Ministerie van Justitie en Veiligheid, 'Memo: State of Play – Microsoft' (17 July 2019).

34. While a detailed analysis of controllership in relation to telemetry data is outside of the scope of this guidance and this submission, the core takeaway is that in relation to access requests, controllers **often conflate customer relationship management systems and broader information management systems**. For example, a great deal of telemetry data is gathered in the video games industry (exactly what online characters did, said, and the like) but will likely not be accessible to customer services, given that it will be held in engineering departments. Similarly, websites and apps gather detailed usage data through invasive approaches such as 'replay' scripts.[16]

35. As it stands, while some organisations have provided such data when they are pushed, it has required data subjects to be aware of the entire structure of the software industry and common practices in order to realise it is likely to be collected. This is unacceptable given the right of access' purpose, which in line with the accountability and transparency principles is (in part) to rectify information asymmetries and allow data subjects to know what controllers know— not to demonstrate what data subjects know about controllers.

36. Controllers must recognise that these datasets fall within the scope of the right of access. Researchers have also pointed to the fact that **privacy departments in these companies are not sufficiently connected or given the authority to talk to developers to ascertain the answer to data subjects' questions**.[17]

Recommendation: Highlight that many vendors (processors and joint controllers) collect telemetry data, which will generally be personal data, and that for companies with data subject-facing systems and interfaces, such as apps or websites, telemetry data must be part of a data access request response to 'all personal data' even if it is not part of e.g. a customer relationship management system.

Recommendation: State that those carrying out data rights within a controller must have the authority and competence to request information and data from other parts of the controller, such as software development, and to dialogue with them as appropriate.

Forms and Time Limits

37. We are glad to see the Commissioner continue to hold the line from the previous Code of Practice that a data controller cannot mandate use of a particular form. We would like to see some small additional clarification, that **the time limit starts from when a valid request is received in any format, not a preferred format or following the provision of information that was already provided in the original request**. We have observed approaches to delay SAR compliance by asking repetitive or

[16] Englehardt and others (n 7).
[17] Ausloos and Dewitte (n 6).

unnecessary questions which have already been clearly addressed in the initial request.

Recommendation: Clarify that the SAR time limits begin from the receipt of a valid SAR, rather that the receipt of an invited particular form.

Verification

38. The Guidance contains good information on verification but it mostly relates to 'typical' data controllers rather than data intensive firms in the digital economy, and/or information society services.

39. **Data controllers often ask for a government issued identification document in situations where it is clearly disproportionate.** In many cases, for example when an individual is seeking data connected to an identifier (eg a cookie ID) and the controller claims no knowledge of the real identity of the data subject, it is unclear what purpose the government ID serves. Moreover, asking for a government ID entails unnecessary risk as data controllers may not have secure systems set up to receive such data, and often in the authors' experience request it through email. Furthermore, in many cases a data subject will be requesting data on the basis that they do not trust the data controller, and wish to consider their options in terms of e.g. objection, erasure or the withdrawal of consent. In these cases, the need to provide sensitive data to the data controller may be unfairly dissuading data subjects from exercising their rights. Some recommendations of national DPAs recommend controllers to request a government ID. **The Guidance should make clear that a government-issued ID should only be required when this is proportionate.** This would also provide reassurance to data controllers who may feel obliged to ask for such information.

40. **Controllers need reassurance that the ICO will not generally seek to take enforcement action against them were they to supply data in response to a fraudulent data access request where they have asked for only reasonable verification.** The guidance should provide this assurance.

41. **Specify that where an individual is 'logged in' to a service, it will generally be disproportionate to require them to provide additional information to verify themselves unless such a requirement is manifestly justified.**

42. **There is a worrying trend for data controllers to design invasive data systems and claim that no amount of verification will ever satisfy them in relation to access requests.** This is particularly common in markets for online tracking, which the ICO has repeatedly described as high-risk and operating largely outside data protection law. In these cases, data controllers are able in practice to single an individual out for, for example, targeting, but they either argue they would never be happy with the level of verification a data subject could provide, or that they deny the data

subject access to their identifier (for example, an identifier hidden in an app on a mobile phone).[18] **The ICO should clearly make a statement that this is not permitted, and that data protection by design places a legal obligation on data controllers to actively design their systems such that access rights (among other data rights) can be fulfilled.**

Recommendation: Specify that a government-issued ID should only be required when it is clearly proportionate.

Recommendation: Specify that where an individual is 'logged in' to a service, it will generally be disproportionate to require them to provide additional information to verify themselves unless such a requirement is manifestly justified.

Recommendation: Specify that the ICO will not normally take enforcement action against data controllers who, despite asking for only reasonable identification, were misled into revealing data to individuals fraudulently claiming to be a different data subject.

Recommendation: Specify that data protection by design requires data controllers to design systems that single individuals out in such a way that access rights can be adhered to. There should be few situations where

Placing Boundaries on 'Manifestly Excessive' Requests

43. The guidance must, in our view, explicitly refute a growing claim that large scale data processing operations can claim an exemption for being 'manifestly excessive' by virtue of their size. There are several reasons for this.

44. **Accepting that a data controller can scale the extent of their data collection and processing per data subject without being subject to a regulatory requirement at least in line with this scaling sufficiently large or complex processing operation sets a dangerous precedent that some data processing activities are 'too big to regulate'.** This logic would mean to say that some processing activities are at such a global scale, and so complex, and producing and capturing so much data about individuals, that they escape the reach of fundamental rights such as the right to access. This seems perverse: the more impactful and the more sizeable the activity, surely the higher the acceptable cost of compliance on the data controller, and the more urgent and pressing the need to provide data subjects with oversight and control rights.

45. Indeed, where such processing implicates a high number of users, this would likely count as 'large scale' processing posing a high risk under the GDPR, and thus has

[18] Michael Veale and others, 'When Data Protection by Design and Data Subject Rights Clash' (2018) 8 International Data Privacy Law 105; Chris Norval and others, 'RECLAIMING Data: Overcoming App Identification Barriers for Exercising Data Protection Rights' in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (UbiComp '18, New York, NY, USA, ACM 2018).

little ground to be manifestly 'unfounded'. Compliance should scale *up* in relation to high risk processing, not down.

46. The ICO should therefore hold firstly that the **manifestly unfounded or excessive** claim relates to the quality of the request in the context of the scale of data processing by the controller, not the burden placed on the data controller. This is in line with existing accepted EDPB guidance, which states (in relation to data portability) that 'the overall system implementation costs should neither be charged to the data subjects, nor be used to justify a refusal to answer portability requests.'[19]

Recommendation: Specify that use of the 'manifestly unfounded or excessive' only relates to the character of the request not the controller-specific burden placed on following it, and that the use of this exemption will be of limited applicability in the context of large-scale processing.

47. In a similar vein, the Guidance must be updated from the Code of Practice to account for the increased prevalence of **large-scale, continuous processing in relation to the rules around 'repetitive' access requests**.

48. Information society services and similar controllership operations typically operate rapidly on large, changing datasets. This is unlike data protection's origins, which implicitly assume reasonably static, unchanging datasets. **Datasets can change in consequential ways in a short period of time, and limitations on access rights should not prevent a data subject from exercising their rights (either at all or without cost) where this is true.**

49. When personal data, and how it is processed, constantly changes, repeatedly exercising data subject rights should not be considered excessive. **Instead, it should be upon such controllers to ensure an automated and easy manner to facilitate the accommodation of those rights. This is in line with data protection by design, and proportionate given that continuously changing datasets imply large scale automated processing, and such automated processing practices can also be purposed towards delivering access rights.**

Recommendation: State that in the context of continuous processing, data controllers should not rely on refusing 'repetitive' requests, because there has been material difference in the data between when requests have been made. Data controllers can continue to only provide data that has actually changed, been inferred or added.

Proposed Example: *A gaming platform runs a dynamic data collection and scoring system which determines individual's visibility to other players. This data is updated every day, and the score is updated accordingly. A data subject makes two requests within a month for this changing data. The data controller is not permitted to refuse the request on the*

[19] Article 29 Working Party (n 14) 15.

basis that it is 'excessive, in particular because of [its] repetitive character', because the data processing operation is of a similar character. Instead of refusal, the data controller must either honour the requests or justify refusal under some other basis. This is proportionate as, in line with the obligation of data protection by design (Article 25), the data controller should be implementing technical and organisational measures to ensure data rights keep pace with data processing, such as providing more regular access to the personal data through, for example, an API or automated data download.

Data Format of Access Requests

50. In older data systems such as that which the Data Protection Act 1998 and the Data Protection Act 1987 anticipated, the number of data points on any given individual was considerably smaller than it often is today. A simple print-out or summary would have sufficed to give the data subject oversight as to the content of the data undergoing processing.[20] **Today, data systems collect such a large number of data points that only a format that allows the data subject to analyse data themselves will allow them to have sufficient oversight over the data processing being undertaken.**

51. Expectations of the data format of access requests can and should be interpreted from the principle of fairness,

52. Firstly, it can and should be understood as part of the principle of fairness that a data controller should not transform data from the machine-readable format they hold it in[21] into a format that makes it more difficult for the data subject to navigate it. Information society services *can only analyse the data they hold about individuals by virtue of its machine-readable nature*. To refuse individuals the same ability exacerbates the informational and power asymmetries that the right of access, and the right of data protection in general, seeks to rebalance.

53. In that context, it is important to recognise that the commonly used document format, PDF, is **not machine-readable, and raises significant difficulties for accessibility and re-use**. Machine readable formats include CSV or JSON or XLSX. In contrast, portable document format, or 'PDF', is a file *designed for printing, not*

[20] This is not to say that many systems have not been considerably complex in relation to subject access rights for many decades, see e.g. Graham Greenleaf and Roger Clarke, 'Database Retrieval Technology and Subject Access Principles' (1984) 16 The Australian Computer Journal.

[21] In its Guidelines on Transparency the A29WP (Article 29 Working Party (n 8) 25) refers to Recital 21 of Directive 2013/37/EU for a definition: 'A document should be considered to be in a machine-readable format if it is in a file format that is structured in such a way that software applications can easily identify, recognise and extract specific data from it. Data encoded in files that are structured in a machine-readable format are machine-readable data. Machine-readable formats can be open or proprietary; they can be formal standards or not. Documents encoded in a file format that limits automatic processing, because the data cannot, or cannot easily, be extracted from them, should not be considered to be in a machine-readable format. Member States should where appropriate encourage the use of open, machine-readable formats.'

for analysis. The EDPB recognised this in accepted guidance on the right to portability, stating that:

> As an example, providing an individual with .pdf versions of an email inbox would not be sufficiently structured. E-mail data must be provided in a format which preserves all the meta-data, to allow the effective re-use of the data. As such, when selecting a data format in which to provide the personal data, the data controller should consider how this format would impact or hinder the individual's right to re-use the data.[22]

54. **PDFs score extremely poorly for individuals who need accessible information online.** Individuals who require or are assisted by accessible information include those with cognitive disabilities, those with vision impairments, those with physical disabilities and those with hearing impairments.[23] A study of 100 blind screen-reader users found that inaccessible PDFs were one of the main causes of frustration while browsing the Web.[24] Accessible PDFs in practice are rarely found, are difficult to create and often require consultants and in-depth planning and expert knowledge.[25] In general, **PDFs are not a tool that lends itself to accessibility across the population.**[26]

55. While other legislation exists which will interplay with the right of access, such as the right to reasonable adjustments in equality legislation, **we are concerned that data controllers are deliberately sabotaging the results of access requests to make them less usable for scrutiny by data subjects.** It is clear from the legislative text that the right of access does not grant a right to machine readable data. However, **we believe it is also clear that the right of access, read in line with the data protection principles, clearly grants a right not to have machine readable data transformed into an inaccessible format to deliberately obscure information, hinder scrutiny, and disadvantage individuals with visual and other impairments.**

Recommendation: Specify that while data controllers who do not hold data in machine-readable format are not obliged to render it machine readable for the purposes of the right of access, data controllers who hold machine readable data are obliged by the principle of

---

[22] Article 29 Working Party (n 14) 14.

[23] cf Gian Wild and Daniel Craddock, 'Are PDFs an Accessible Solution?' in *Computers Helping People with Special Needs* (Lecture Notes in Computer Science, Klaus Miesenberger and others eds, Springer International Publishing 2016) 355.

[24] Jonathan Lazar and others, 'What Frustrates Screen Reader Users on the Web: A Study of 100 Blind Users' (2007) 22 International Journal of Human–Computer Interaction 247.

[25] Erin Brady and others, 'Creating Accessible PDFs for Conference Proceedings' in *Proceedings of the 12th Web for All Conference* (W4A '15, New York, NY, USA, ACM 2015).

[26] ibid.

fairness *not* to transform it into a non-machine readable format, such as a PDF, as this hinders its scrutiny by data subjects.