



**HAL**  
open science

# GDPR compliant guidelines for processing personal data in legal documents

Noora Arajärvi, Livia Holden, Anna Ziliotto, Joshua Isaac Bishay

## ► To cite this version:

Noora Arajärvi, Livia Holden, Anna Ziliotto, Joshua Isaac Bishay. GDPR compliant guidelines for processing personal data in legal documents. 2022. hal-03674949

**HAL Id: hal-03674949**

**<https://hal.science/hal-03674949v1>**

Submitted on 21 May 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

This document is one of the primary outputs of EURO-EXPERT, European Research Council funded project led by Livia Holden, and it is protected by copyright laws. Please cite as follows: Noora Arajärvi and Livia Holden (with the assistance of Anna Ziliotto and Joshua Bishay) 2021 “GDPR compliant guidelines for processing personal data in legal documents”, available at <https://culturalexpertise.net/wp-content/uploads/2022/01/gdprcompliantguidelinesforlegaldocuments.pdf>



## **GDPR compliant guidelines for processing personal data in legal documents**

Noora Arajärvi and Livia Holden (with the assistance of Anna Ziliotto and Joshua Bishay)

1. Aim and scope of the guidelines
2. What is anonymisation and why to anonymise?
3. What to anonymise?
4. What not to anonymise?
5. Special categories
6. How to anonymise?
7. Challenges and tips

---

Annex 1: Identifying legal persons for anonymisation

Annex 2: Practical steps for redacting information related to legal persons

---

### **1. Aim and scope of the guidelines**

CULTEXP, the main output created by EURO-EXPERT and the first multilingual and cross-jurisdictional database on cultural expertise, contains judgements and expert reports that have been collected from the following sources: OPEN ACCESS databases, law court archives, and experts. The law on the processing of personal data in the European Union is set out in the General Data Protection Regulation 2016/679 (the GDPR) and national legislation. At the time of writing these guidelines, interpretations and practices regarding the processing of personal data in legal documents vary greatly. At EURO-EXPERT, we have developed specific GDPR compliant guidelines for the treatment of personal data contained in the documents archived on CULTEXP. In developing these guidelines, we have referred to the GDPR, to other regulations and case law, and to the expert opinion of Prof. Jogleux (European University Cyprus) who was appointed by EURO-EXPERT to address the purpose of CULTEXP and the guiding principles of personal data processing in legal documents.

### **2. What is anonymisation, and why to anonymise?**

The concept of personal data under the GDPR<sup>1</sup> is very broad. Personal data is not only “direct personal data” that allows immediate identification of the data subject (such as the name or title, e.g. Prime Minister of France) but also “indirect personal data” (e.g. 45-year old Greek female professor living in Cambridge). The concept of indirect personal data embraces any data that could lead to identification using reasonable means.

---

<sup>1</sup> Complete Guide to GDPR Compliance: <https://gdpr.eu/>.

Anonymisation is a technique of data processing that can be applied to personal data to achieve irreversible de-identification. One of the advantages of anonymisation is that it allows for research that would otherwise not be possible due to privacy concerns. This means removing direct identifiers such as names, birth dates and addresses, although this alone may not be sufficient to avoid the identification of individuals. Therefore, each case must be considered in its entirety to ensure that no individuals are identifiable.

Effective anonymisation prevents singling out an individual in a dataset by linking several records within the same dataset (or between several separate datasets) or inferring relevant information from such dataset(s). The aim of anonymisation is to ensure that the data no longer relates to identifiable persons, which means it is no longer considered “personal data”. Therefore, once a dataset is properly anonymised and individuals are no longer identifiable, the data falls outside the scope of data protection rules under the GDPR. In 2018, the Court of Justice of the European Union introduced measures it was adopting to comply with the GDPR in the publication of its decisions.<sup>2</sup> National legislation regarding privacy and data management is also relevant in this regard.

Pseudonymisation consists of replacing one attribute with another. It is defined in the GDPR as “*the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.*”<sup>3</sup> Most EU legal systems use a two-step technique to achieve the most effective outcome: First, the names of the parties to the litigation are changed to their initials. Then, personal identifiers such as ID numbers or addresses are replaced with “x” characters. Pseudonymisation, however, may require further redaction to comply with the GDPR. Even when the name of a party is completely replaced with initials, if the decision describes a situation that is unique or that is very well-known, the decision may still contain personal data. Similarly, if, by using techniques of data mining, users could find cross-references leading to specific identifiers of persons involved, the data may still be within the scope of the GDPR and national legislation. It is important to note that member states have implemented, interpreted, and applied the GDPR somewhat differently across the EU. Hence, an *ad hoc* assessment of the balance between the risk of disclosing personal data and a legitimate purpose of personal data processing may be required in particular cases.

Generalisation is another anonymisation technique. This approach consists of generalising or diluting the attributes of data subjects by modifying the relevant scale or order of magnitude (e.g. using a region rather than a city, a month rather than a week, etc.). Whilst generalisation can be effective to prevent the singling out of individuals, and may be useful as a supplemental tool, it does not allow for effective anonymisation in all cases.

Processing personal data for research purposes occupies a privileged position within the GDPR.<sup>4</sup> CULTEXP (re-)publishes caselaw according to a legitimate interest of facilitating access to justice on topics related to cultural expertise.<sup>5</sup> In the phase of data-treatment, the EURO-EXPERT team removed **all personal data from their data sets**,

---

<sup>2</sup> Court of Justice of the European Union, ‘From 1 July 2018, requests for preliminary rulings involving natural persons will be anonymised’, Press release No 96/18, Luxembourg, 29 June 2018, <https://curia.europa.eu/jcms/upload/docs/application/pdf/2018-06/cp180096en.pdf>.

<sup>3</sup> Article 4 (5) GDPR.

<sup>4</sup> Art. 89 GDPR on the safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes. Also, G Maldoff, ‘How GDPR changes the rules for research’ (2016): <https://iapp.org/news/a/how-gdpr-changes-the-rules-for-research>.

<sup>5</sup> With precise, specific, and detailed legitimate interest, in the sense described in the Working Party Opinion 06/2014, “Notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC”, accessible at [https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217\\_en.pdf](https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf). See also Art. 89 GDPR.

unless a clearly specified exception applies. EURO-EXPERT has adopted the following combination of anonymisation and pseudonymisation, which is explained in further detail below:

- 1) The names of the parties are reduced to initials or have been redacted or removed.
- 2) All personal identifiers such as ID numbers, addresses, social security numbers and so on are removed or redacted.

### 3. What to anonymise?

EURO-EXPERT requires anonymisation of:

1. Names of **all** living individuals unless otherwise specified: Defendants, plaintiffs, claimants, applicants, appellants, respondents, victims, witnesses, other parties, lawyers, experts and consultants.
2. Names of places: Addresses (home address and other); specific locations (villages, towns, and cities, depending on the risk that individuals might be identifiable with this information – consider generalisation); specific places (hotels, shopping centres, shops, farms, churches, accommodation, hospitals, buildings or monuments, boats, etc).
3. Dates: Birth dates (consider generalisation: a year instead of a date); other dates or years which could make a person identifiable (dates suggesting specific and recognisable events in the life of subject, such as the date of marriage, birth dates of children); dates of other decisions, orders and notifications (asylum application, decision by an administrative authority, etc).
  - a. In some cases, redacting only part of the date may be sufficient if the timeline is relevant for understanding the case. Consider whether the month/year may be retained, for instance, in cases where the age of the individual is relevant. For example:

“It is claimed they were born 30 Jan 2021, whereas other documents say they were born 12 Dec 2021”

There are two options; the first option provides more information:



It is claimed they were born [Redacted] Jan 2021, whereas other documents say they were born [Redacted] Dec 2021


OR



It is claimed they were born [Redacted] 2021, whereas other documents say they were born [Redacted] 2021

- b. In other cases, the specific dates can be removed without losing key information. For example:

The asylum seeker claims: "sono in Italia dal 30 agosto 2015 e vengo dal Senegal, Kolda. Ho 28 anni, non sono sposato e non ho figli. Ho lasciato il Senegal nel maggio 2013. Sono andato in Mali, poi in Niger, in Libia dove sono stato nel 2014 per un anno."


 sono in Italia dal [removed] e vengo dal Senegal, [removed]. Ho 28 anni, non sono sposato e non ho figli. Ho lasciato il Senegal nel [removed]. Sono andato in Mali, poi in Niger, in Libia dove sono stato nel [removed] per un anno.]

4. Numbers: National identification numbers, passport numbers, fiscal codes, social security numbers, bank accounts, vehicle registration numbers, identification numbers (e.g. personal identification under the asylum procedure), police report numbers (codes of recordings, codes for intercepts), telephone numbers, postal codes, etc, if they provide information that could lead to the identification of individuals.
5. Signatures: In some jurisdictions case files include signatures. Note also signatures written in the margins of the documents (e.g. Italy). Signatures of judges do not need to be redacted if the names of judges are unredacted. Many expert reports include signatures, which should be removed or redacted.
6. Brands, labels, and corporate logos (including those of legal database services) if they provide information that could lead to the identification of individuals.
  - a. In a case of terrorism, the brand of the shirt of the accused was redacted because the eyewitness identified the accused by the branding of their clothes.
  - b. In the UK cases, many decisions were downloaded from Westlaw, and their logo has been redacted.
7. In some cases, names of private associations, institutions, and commercial enterprises are also redacted if those can be linked to and used to identify a natural person. This is addressed in detail in Annex 1 on legal persons and anonymisation. See also Annex 2 on processing decisions related to legal persons. Consider the context, scope and level of detail: the CEO of a named small local company is identifiable, whereas a salesperson working for Carrefour or Tesco in a large city may not be.
  - a. If names of companies are redacted, “nesting” corporate entities may be necessary to retain readability of the case. For example:

“Marvel Studios is a subsidiary of Disney Corp, which purchased it from Marvel Entertainment Group.”

 [Corp 1.1] is a subsidiary of [Corp 1], which purchased it from [Corp 2]

8. In some cases, the case numbers have been redacted. If there is no imminent risk of or access to sensitive information via the case number, this does not have to be redacted. A balancing assessment should be made between the legitimate purpose of the database and the indirect risk of disclosure of personal data; note that the case number might be disclosed for the purpose of legal citations. In some jurisdictions, anyone can order the case from a court with a case number and could access the personal data that way. In the Italian context, the case numbers of decisions of the Commissione Territoriale reveal the social security number of the applicant, which should be removed or redacted.
  - a. In other Italian judgments, some cases are identified as follows:  
"Cass. Sez. 6<sup>^</sup>, 20 ottobre 1999, Bajrami" (or with the name of judge).

 "Cass. Sez. 6<sup>^</sup>, 20 ottobre 1999, B[removed]"

#### 4. What not to anonymise?

Do not remove or redact the following, unless the data can lead to the identification of individuals:

1. Names of judges and court officials, unless country legislation or practice dictates otherwise, or where revealing the name of the judge may put them at a risk of reprisal. For further discussion, see the next section on special categories.
2. Names of deceased persons, historical figures, famous persons, and public figures. However, if a public figure is a party to a case, they may have a right to privacy; consider also issues such as libel suits.
  - a. In some cases, well-known public figures may have been indirectly involved in circumstances surrounding the case. In such instances, consider if it may be useful to not only redact their name but also their role and title. For example:

“Emmanuel Macron was president of France when he signed the contract with John Doe.”



[EM] was [Redacted] when he signed the contract with [JD]

3. Names of criminal organizations (Cosa Nostra, Islamic State, Black AXE, etc).
4. Names of provinces or states, unless such information could lead to the identification of the person(s) – see for example:

“asylum seeker from Hashela district of Keren in Eritrea living in the small town of Grândola in Portugal”

 “asylum seeker from [Redacted] district of [Redacted] in Eritrea living in the small town of [Redacted] in Portugal”

5. URLs.
6. Bibliographical references (authors, titles).
7. Public and international associations or organisations (United Nations, Red Cross, World Health Organization, etc).
8. Names and other data related to legal persons (companies, associations, etc.) when those cannot be linked to an individual and do not lead to the identification of a natural person. Note that, under the GDPR, only natural persons have personal data. This is addressed in detail in Annex 1.
9. Legal precedents that contain the names of the parties (unless these lead to identification of the individuals involved in the case that you are anonymising).
  - a. Consider redacting names when the court refers to names of individuals in discussing the facts of the precedent. For example:

“In this case of Appellant we rely on the *Parker* case. In that case Mr Parker was driving down the street when he got in an altercation with the driver of another vehicle.”



In this case of Appellant we rely on the *Parker* case. In that case [Redacted] was driving down the street when he got in an altercation with the driver of another vehicle.

## 5. Special categories

As a general principle, the GDPR applies to all personal data. However, national legislation and practices on the implementation of the GDPR vary significantly. At EURO-EXPERT, we have debated on whether to anonymise the names of the judges. To the best of our knowledge, no binding interpretation on the application of GDPR to legal documents has been issued to date and available guidelines do not provide instructions on this specific matter. Article 6 (1) (f) GDPR entails an analysis of proportionality: regarding the publication of the names of court staff, and especially judges, the legitimate interest of the publication prevails over the risk posed to their privacy. The publication of the names of judges contributes to democratic debate and may be protected by article 85 of the GDPR. The principle of transparency in the judicial process and the principle of open justice as enshrined in article 47 of the EU Charter of Fundamental Rights support the publication of the names of judges to foster the effective independence and impartiality of the courts. It has been noted that one of the main reasons for publishing court decisions at the national level is the public scrutiny of the judge.<sup>6</sup> Further, it has been observed that “in most countries the names of professionals involved [in a court case] are not rendered anonymous [...]”.<sup>7</sup> According to the expert opinion received by Prof. Jougleux (European University Cyprus), there is no explicit reference in the EU legislation to the names of judges and lawyers in court decisions; furthermore, the mention of the names of the judges and the lawyers does not generally present privacy risks and should therefore not be included in the anonymisation mechanism. Hence, as a general rule, the names of people involved in a court case in their professional capacity are not anonymised in CULTEXP; for example judges, clerks, lawyers, bailiffs, court experts, interpreters and custodians.<sup>8</sup> Exceptions to this may include specific country practices – the respective Data Protection Authorities of different EURO-EXPERT countries have varying opinions and practices on what information should be included in the public court files – and in situations presenting a high risk of reprisals (e.g. terrorism and organised crimes cases).<sup>9</sup>

As country practices vary and it is expected that more detailed guidelines will be developed, EURO-EXPERT endorses a cautious approach overall. EURO-EXPERT abides by the national legislations and regional practices that require the redaction of the judges and court staff’s personal data, constantly monitors the development of the implementation of the GDPR and will adapt to future changes.

At this stage, EURO-EXPERT **anonymises or pseudonymises the personal data on lawyers and experts involved in a court case**. As the guidelines for the processing of personal data in the publication of judgements in legal databases evolve, EURO-EXPERT may review and relax this precautionary approach. The EURO-EXPERT team is advised to also use their expert knowledge of the context and national legislation and practices in combination with GDPR regulations and always to err on the side of caution if in doubt.

---

<sup>6</sup> Michal Bobek, ‘Data protection, anonymity and courts’, 26(2) *Maastricht Journal of European and Comparative Law* (2019) 183, <https://journals.sagepub.com/doi/pdf/10.1177/1023263X19851628>.

<sup>7</sup> Marc van Opijnen, ‘Court Decisions on the Internet: Development of a Legal Framework in Europe’, 24(2) *Journal of Law, Information and Science* (2016) 26. <https://www.austlii.edu.au/au/journals/JLLawInfoSci/2016/3.html#fn66>.

<sup>8</sup> M van Opijnen *et al.*, ‘On-line Publication of Court Decisions in the EU: Report of the Policy Group of the Project ‘Building on the European Case Law Identifier’’, 15 February 2017, <https://www.bo-ecli.eu/uploads/deliverables/Deliverable%20WSo-D1.pdf>, p. 24

<sup>9</sup> T Allard, L. Béziaud & S. Gambis, ‘Online publication of court records: circumventing the privacy-transparency trade-off’, 3 Jul 2020, <https://arxiv.org/abs/2007.01688>.

## 6. How to anonymise?

1. Keep two folders on the Shared Drive: one for the original documents (for reference and if there is a need to check details, but this folder is to be deleted at the end) and one for the anonymised documents. Avoid keeping any original documents in locations (physical or digital) which could be accessed by others or could pose any risk of a data breach.
2. When saving the anonymised document, make sure the file name does not contain any names or other personal data. The same applies to its 'document properties' or metadata.
3. Depending on your dataset, you may consider different options for anonymising or pseudonymising the names of individuals:
  - a. Use initials representing the first name and surname.
  - b. If initials might reveal the identity of the person, use letters or numbers but try to be consistent throughout the case, as long as the letter or number will not reveal the identity:
    - i. A1, A2 etc for claimants/plaintiffs and B1, B2 etc for respondents/defendants)
    - ii. For experts, use E1, E2, etc.
    - iii. For witnesses, use W1, W2, etc.
    - iv. For judges, use J1, J2, etc.
    - v. For lawyers, counsel, attorneys and other legal representatives of parties, use L1, L2, etc.
  - b. If there is still a risk of identifying individuals, use -- or xx for names of all or some of the individuals involved in the case.
4. If the document makes a passing reference to any person, you can redact, or use “—” or “xx”. These can be used when the role of that individual is irrelevant to the narrative as a whole.
5. When going through the case, you may find it useful to keep a pen and paper at hand to write down your “code”, e.g. “Dr Carol Smith = E1”. Make sure to properly destroy these notes afterwards.
6. In MS Word, under “Edit”, you will find “Find” under which there is the function “Replace”. In some versions of MS Word, you will find this function under "Home" > "Editing". Use this with caution. See below for details.
7. If you work with PDF documents, using Adobe may be the only secure option. Remember to sanitise the document at the end!
8. Do not use any online PDF-Word converters for documents containing personal data or sensitive information. Adobe Pro includes a PDF-Word converter. If you have any problems, contact the EURO-EXPERT team.

## 7. Challenges and tips

Removing direct identifiers is not enough to ensure that identification of the data subject is no longer possible; pseudonymised data is not the same as anonymised data!

Simply altering the main identifier (usually the name) does not prevent someone from identifying a data subject if quasi-identifiers (such as places, dates, age and so on) remain in the dataset. Extra steps should be taken, such as removing and generalising attributes or deleting the original data, or at least bringing them to a highly aggregated level.



- Example: if a person is described as ‘a man’, the anonymity set size is three and a half billion, but if he is described as ‘a middle-aged Dutchman with a beard’ the anonymity set size is reduced to, say, half a million. If he is described as ‘a middle-aged Dutchman with a beard who lives near Cambridge’ the anonymity set size is only three or four and consequently allows for a high probability of identification.

Make sure you:

- Read and understand the case and the roles of persons involved.
- Clarify whether the data will be anonymised (the link to the data subject will be destroyed) or pseudonymised (the data could be reversible).
- Do not rely on the “find-replace” function:
  - Documents may be partly corrupt and not all references are identifiable by the search function.
  - Documents may contain inaccuracies, misprints or mistakes.
  - The same name, place, or other identifier may appear with several spellings in the same document (e.g. Mohammed, Mohamed, Muhammad, Muhammed etc).
  - In certain languages (e.g. Finnish) suffixes may alter the core of the word, rendering it undiscoverable by the search function (e.g. Mäki - Mäen).
  - The find-replace function can be a useful supplemental tool in the final check of the document to see if any identifiers remain.
- Remember to check the case name/citation for any identifiers!
- Remember to check also footnotes for any personal data, if applicable.
- If some data will not be anonymised, explain why you cannot anonymise the data.
- If the data will be coded, describe the coding system, and who will have access to it; and confirm that it cannot be traced back to individuals unless essential for the study.
- Assess whether the remaining data combined with additional information might permit the identification of an individual.
- Prepare a short statement disclosing your anonymisation technique or the mix of techniques that you have used, and highlight any caveats, challenges, or problems. Doing so will help to show that your work is conscientious and reflective and will serve to maintain the quality and legal compliance of the outcomes of the project.

## Annex 1: Identifying legal persons for anonymisation

### 1. Introduction

There are several reasons for anonymising or redacting the name of a legal person, such as a corporation. The two most common reasons are (1) to maintain readability, without having to sacrifice the strict standards of GDPR compliance in EURO-EXPERT, and (2) to remove any data which could reveal information about unrelated parties.

The need to anonymise or redact the name of a legal person arises most commonly in proceedings where the actions of natural persons are key to litigation between legal persons, for example, when the CEO of a company is accused of defrauding another company's corporate officers. In these situations, there are two options: (1) to anonymise or redact the natural person's name and relationship with the legal person, or (2) to anonymise or redact the natural person's name and the legal person's name.

### 2. Maintaining comprehension and readability

In Excerpt 2.1, it is clear that the natural person is the founder of a corporate entity. The two options are shown below to fully anonymise the paragraph:

*Excerpt 2.1:*

2 The First Defendant, VAK ("VAK "D"), is a Russian citizen and the founder of the JFC Group. The Second Defendant, JFC Group Holding (BVI) Limited

In Excerpt 2.1(a), the natural person's data redacted:

*Excerpt 2.1(a):*

2 The First Defendant, VAK ("VAK "D"), is a Russian citizen and the founder of the JFC Group. The Second Defendant, JFC Group Holding (BVI) Limited

In Excerpt 2.1(b), the legal person's data redacted:

*Excerpt 2.1(b):*

2 The First Defendant, VAK ("VAK "D"), is a Russian citizen and the founder of the JFC Group. The Second Defendant, JFC Group Holding (BVI) Limited

In excerpt 2.1(a) the relationship that VAK has with the company is lost, although we know the name of the company. However, with Excerpt 2.1(b) the pseudonymised name of the legal person is "Corp1" (Corporate 1), so we know that VAK is the founder of "Corp1" and more information is preserved to improve readability. Furthermore (in Excerpt 2.1(b)), leaving "the founder" unredacted does not risk disclosure of personal data but maintains the importance of this person's position in the company as a measure of their decision-making power, or to contextualise the later evidence discussed in the case. For example, with excerpt 2.2:

*Excerpt 2.2:*

e. At the suggestion of Mrs YZ, Mr VAK agreed that I should take over as General Director of JFC Russia from April 2011 so that Mrs YZ could concentrate her efforts on finding alternative sources of funding.

In Excerpt 2.2(a) below, the testimony of the witness would be useless unless one is allowed to know that YZ, who was CEO at the time, convinced the founder of "Corp1" that the witness should become GD of Corp1's subsidiary "Corp1.1" as below would maintain:

Excerpt 2.2(a):

e. At the suggestion of Mrs YZ, Mr VAK agreed that I should take over as General Director of JFC Russia from April 2011 so that Mrs YZ could concentrate her efforts on finding alternative sources of funding.

Another example (Excerpt 2.3, below) of this arises in relation to real property (buildings, land, apartments, etc.). In the example below, there is a question of whether Loc3 (Location 3) is owned by Corp10.1 or VA. By redacting the precise location and the company name, this context can be maintained so that the reader understands the dispute between Corp10's subsidiary (Corp10.1) and VA.

Excerpt 2.3:

(2) A residential development at Solnechnoye, to the north of St Petersburg and near Loc3 on the Gulf of Finland, owned by Corp10.1, a subsidiary of Corp10 which DrVA says was owned by him;

### 3. Protecting the identity of third parties uninvolved in litigation

In Excerpt 3.1, disclosing the legal entity's name would reveal the involvement of a natural person who is entirely uninvolved in the case:

Excerpt 3.1

9 Between June and early November 2016, Corp4 was engaged by Corp5 ("Corp5"), a consultancy based in Washington DC, providing research, strategic intelligence and due diligence services to clients. Mr CS and Corp4 had developed a prior working relationship with Corp5 over a number of years. Corp5 engaged Corp4 to prepare a series of confidential memoranda based on intelligence concerning

DJT. The parties describe these memoranda as the and

In Excerpt 3.1, redacting the names of two companies is necessary to protect the identity of CS, but it also conceals further information about individual DJT, who is not a party or witness in the case. By redacting the names of the companies, we do not have to redact "had developed a prior working relationship with Corp5 over a number of years".

## Annex 2: Practical steps for redacting information related to legal persons

When anonymising and assigning pseudonyms to cases involving legal persons (corporations), it may become necessary to redact some data of the legal persons to protect the identity of natural persons (individuals). This is especially important when looking at cases examining the action and behaviour of named corporate officers (CEOs, CFOs, CTOs, etc.) or named corporate employees. Unfortunately, these types of cases, especially involving accusations of “white collar crimes” (such as embezzlement, tax evasion, and fraud) often include complex and opaque webs of corporations, shell companies, offshore accounts, and natural persons. As demonstrated below, in order to maintain readability of the document, this complex web should be understood by the redactor before they begin. This can be achieved by: (1) mapping out the corporate structures, (2) identifying outside organisations, (3) identifying properties owned by companies and (4) identifying extraneous information about the company which may not relate to the case at hand but can lead to the identification of the company (and by extension to the involved natural persons).

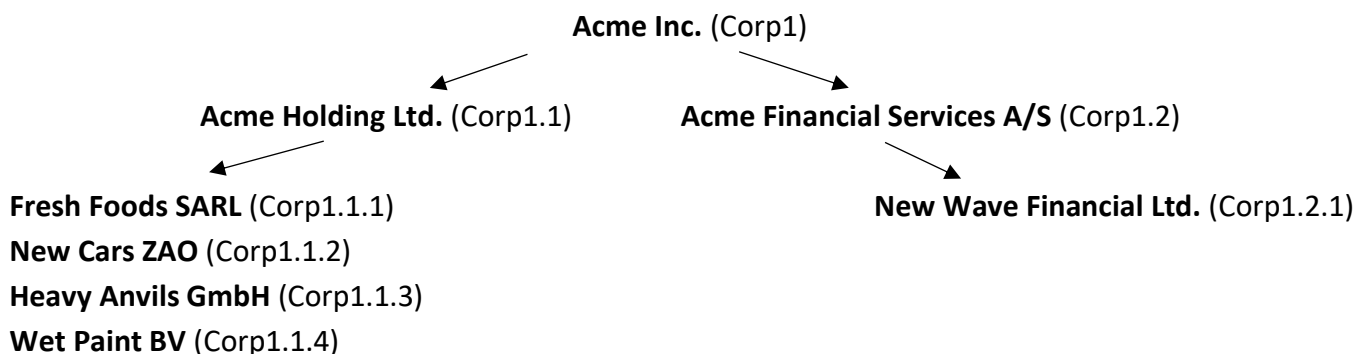
### 1. Mapping out the corporate structure

To maintain readability, it is important to make sure the pseudonyms reflect the relationships between parties according to the corporate structure, and to have a clear plan on how the information about these relationships can be incorporated into the pseudonymisation and to ensure not to double number them. The easiest and most intuitive way to do this, without using too much space, is to label the highest-level company with a whole number (Corp1, Corp2, Corp3, etc.), and as you move down within a single corporate structure add decimals to reflect where the company in question lies on the corporate map (Corp1.1 being the first company encountered which is a subsidiary of Corp1). Using the fictional example below, this can look as follows:

*Example 1.1:*

9. Mr. John Doe was, at all relevant times, the CEO and primary shareholder (80%) in Acme Holding Ltd. (“Acme Holding”), a subsidiary of Acme Inc. (“Acme”). Acme Holding wholly owns a wide variety of companies such as Fresh Foods (France) SARL, New Cars (Russia) ZAO, Heavy Anvils (Austria) GmbH and Wet Paint BV in the Netherlands. Acme also owns the Acme Financial Services A/S (“Acme Financial”), with Ms. Jane Roe as its CEO and its principal offices located in Copenhagen, which has the primary function of providing financial assistance to subsidiaries of Acme Holding. Acme Financial is 70% owner of New Wave Financial, based in London.

The map for Example 1.1 would look as follows:



After mapping out the corporate structure, it is possible to replace the names (and the natural persons involved) with the assigned pseudonyms. Be sure to leave names of countries, when possible, as this

helps the reader to follow judicial logic in some situations, especially when dealing with offshore shell companies as will be seen in a later example. This should look as follows:

*Example 1.1(a)*

9. Mr. JD was, at all relevant times, the CEO and primary shareholder (80%) in Corp1.1 ("Corp1.1"), a subsidiary of Corp1 ("Corp1"). Corp1.1 wholly owns a wide variety of companies such as Corp1.1.1 (France), Corp1.1.2 (Russia), Corp1.1.3 (Austria) and Corp1.1.4 in the Netherlands. Corp1 also owns the Corp1.2 ("Corp1.2"), with Ms. JR as its CEO and its principal offices located in Copenhagen, which has the primary function of providing financial assistance to subsidiaries of Corp1.1. Corp1.2 is 70% owner of Corp1.2.1, based in London.

Unlike in Example 1.1 and 1.1(a) often the relationship between companies is not apparent until later on in the decision, as demonstrated in Example 1.2 below:

*Example 1.2*

9. Mr. John Doe was, at all relevant times, the CEO and primary shareholder (80%) in Acme Holding Ltd. ("Acme Holding").

....

102. Mr. John Doe met on 31 February 2020 with Ms. Bertrand Roe, CEO of Acme Inc., the parent company of Acme Holding.

Again, when redacting the above example, if one had not made it to paragraph 102 yet, it is easy to confuse Acme Holding as Corp1, and find out later that Acme Inc. is actually the parent company. This would require one to go back and correct the mistake. To avoid this, check the entire document first to see if the relationship is described. Thus, when done correctly it should look like Example 1.2(a) below.

*Example 1.2(a)*

9. Mr. JD was, at all relevant times, the CEO and primary shareholder (80%) in Corp1.1 ("Corp1.1").

....

102. Mr. JD met on 31 February 2020 with Ms. BD, CEO of Corp1, the parent company of Corp1.1.

When a company's relationship to other companies is in question, or the source of litigation, it can make deciding where to place them in the corporate map difficult. Such an example can be seen below in Example 1.3:

*Example 1.3*

35. 30 million Euros was transferred between Roi Mug SARL and Lamp Manufacturer Inc., a subsidiary of Lamp Inc., on 31 April 2020. The Complainant argues that Roi Mug SARL is under the direct control of Mr. John Doe's holding company Glassware Holding Ltd. However, Mr. John Doe argues that he has never heard of Roi Mug, and that he has no contacts with anyone associated with it whatsoever.

In Example 1.3, the relationship between Roi Mug SARL and Glassware Holding Ltd. is unclear, and may be subject to judicial review to determine whether it is actually related to Glassware holding Ltd. In this case, refer to Roi Mug with a letter, rather than a number. In Example 1.3(a) I referred to Roi Mug SARL as Corp X, but if more than one corporation is involved, consider pseudonymising in alphabetical order

(CorpA, CorpB, CorpC). Furthermore, if more than 26 companies (because there are 26 letters in the English alphabet) are involved, double letters may be necessary (CorpAA, CorpAB, CorpAC).

*Example 1.3(a)*

35. 30 million Euros was transferred between CorpX and Corp2.1, a subsidiary of Corp2, on 31 April 2020. The Complainant argues that CorpX is under the direct control of Mr. JD's holding company Corp1.1. However, Mr. JD argues that he has never heard of CorpX, and that he has no contacts with anyone associated with it whatsoever.

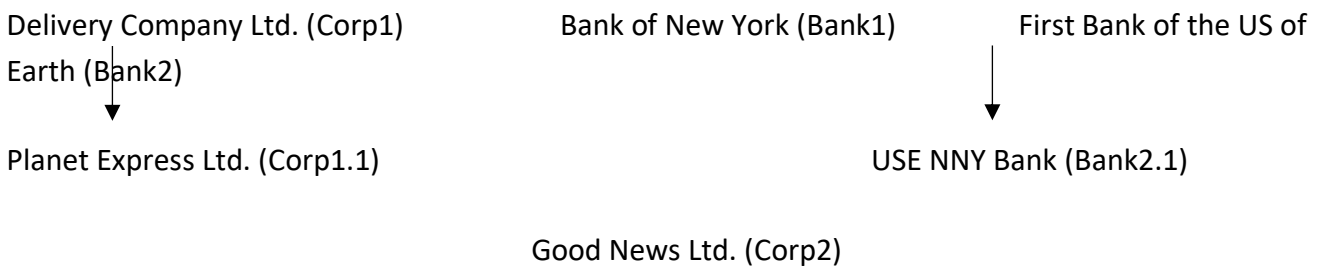
**2. Outside Corporations and Groups**

When a company is not related to another company, simply assign it the next number (for example if you have already assigned Corp1, it would be Corp2, then Corp3). In more complex cases, you can end up easily with over 50 different and unrelated companies. It is important, however, to distinguish between banks (Use Bank1, Bank2, etc.) and companies – also keeping in mind the necessary corporate maps which may be needed. For example, in Example 2.1 below:

*Example 2.1*

41. In the year 2000, Delivery Company Ltd.'s ("Delivery Company") CFO Ms. Jane Doe reached out to Bank of New York and First Bank of the United States of Earth's subsidiary in New York, USE NNY Bank on behalf of Delivery Company's subsidiary Planet Express Inc. Delivery Company requested a loan to engage in a building project with Good News Ltd. to build a new warehouse.

In Example 2.1, the following maps would be created:



In Example 2.1, Good News is not affiliated with the Delivery Company family of companies, nor is its relation in question, it is only potentially entering into an agreement with Delivery Company's subsidiary, thus it would be Corp2. Secondly, the First Bank of the United States of Earth (Bank2) has a subsidiary too, USE NNY Bank and so this should be given the pseudonym Bank2.1. When we apply these redactions and their associated pseudonyms, the information of who is a Bank, who is a company, and which companies are related is maintained. As we see in Example 2.1(a) this would look like this:

*Example 2.1(a)*

41. In the year 2000, Corp1's ("Corp1") CFO Ms. JD reached out to Bank1 and Bank2's subsidiary in New York, Bank2.1 on behalf of Corp1's subsidiary Corp1.1, Corp1 requested a loan to engage in a building project with Corp2 to build a new warehouse.

This distinction between banks and companies is useful to the reader, even though a bank is technically a corporation. If this distinction is not made, it could confuse the reader into thinking that a company gave money to another company, as opposed to the very different legal implications of a loan from a bank.

### 3. Identifying properties owned by a company

When a company owns a property, especially real property, it could lead to identification of the company – especially when the property is well known. As redactors, we should not assume what properties are well known to the local population or even internationally and redact all properties which are associated with the company.

#### Example 3.1

41. Book Sellers BV's CEO Ms. Jane Doe met in Amsterdam with representatives from Storage Solutions BV to permanently purchase the warehouse at Zonweg 21 in Den Haag near the Voorburg train station from Storage Solutions to use to store their imported books for sale throughout the Netherlands.

...

100. Book Sellers, after a series of successful sales raised enough capital to purchase a 40% interest in the Den Haag Municipal Library. The Library was not far from the warehouse.

There must be a distinction here between the three types of locations used: (1) the cities, places and countries where events and properties are located such as Amsterdam, Den Haag, the Netherlands and Voorburg Station; (2) exact addresses such as Zonweg 21; and (3) buildings such as Den Haag Municipal Library or the warehouse. Only the latter two need to be redacted, as the location of meetings and countries in which the property is located can have important legal repercussions, and generally do not risk disclosure of personal data. This is also the case in Examples 1.1 and 1.1(a), which name the country where each company is located and in Examples 2.1 and 2.1(a) where New York is named. When redacted, Example 3.1 should look like this:

#### Example 3.1(a)

41. Corp1 's CEO Ms. JD met in Amsterdam with representatives from Corp2 to permanently purchase Location1 in Den Haag near the Voorburg train station from Corp2 to use to store their imported books for sale throughout the Netherlands.

...

100. Corp1 , after a series of successful sales raised enough capital to purchase a 40% interest in the Location2 . Location2 was not far from Location1 .

In Example 3.1(a), notice that the entirety of “the warehouse at Zonweg 21” in paragraph 41, and similarly in paragraph 100 “the library” and “the warehouse” are redacted. This is because, as a redactor who is unfamiliar with the local geography, it would be too time-consuming to look up the details and the identifiability of the property, thus it is safer to redact. Furthermore, as with “the warehouse” in paragraph 100, for consistency and clarity one should continue to label “the warehouse” as Location1 and “the library” as Location2.

Also in Example 3.1(a), although one could redact “near Voorburg train station”, the previously redacted information is sufficient to protect the identity of Jane Doe (the person whose identity we are trying to protect). The goal is to reach the maximum protection of personal data with the minimum number of redactions.

Another example of a name of a property needing to be redacted is when naming a chattel (a tangible thing) can lead to the identification of natural persons. Chattel can be (non-exhaustively) ships, train lines, airplanes or products. This is demonstrated below in Example 3.2:

### Example 3.2

42. Planet Express Inc. agreed to purchase the USS Enterprise, an Ocean-Class ship, from Cylon SARL to connect Planet Express' Spice Line train in Bordeaux with their Tribble factory in San Francisco to sell Tribbles throughout the EU.

....

2001. After Planet Express' company grew, they engaged with the Bank of Tercon Prime to become the first purchaser of three Daedalus airplanes from The Guild Inc. to supplement their subsidiary's Airbus fleet.

One must use their judgement to determine if the information can lead to the identification of a legal or natural person. In the above example, as ships are given unique names which make them easy to identify, USS Enterprise should be redacted. Additionally, "Spice Line" is easily identifiable as a specific train line. Finally, a "Tribble", a fictional product in the above example, should be redacted as it may clearly connect the tribbles to Planet Express Inc.

### Example 3.2(a)

42. Corp1 agreed to purchase Ship1, an Ocean-Class ship, from Corp2 to connect Corp1, Train1 in Bordeaux with their Product1 factory in San Francisco to sell Product1 throughout the EU.

....

2001. After Corp1's company grew, they engaged with Bank1 to become the first purchaser of three Plane1 airplanes from Corp3 to supplement their subsidiary's Airbus fleet.

In Example 3.2(a), each form of property was generalised. For example: Tribbles are products, the USS Enterprise is a ship, the Spice Line is a train line and the Daedalus is an airplane. Generalisation makes it easiest for the reader to follow the decision, especially when the products may only be mentioned hundreds of pages apart. Although named, it may not be necessary to redact Airbus; There are hundreds of companies which own Airbus planes, and thus it is not necessary to pseudonymise it into "Plane2".

## 4. Identifying extraneous information about the company

Sometimes, there may be information about a company which does not fit into the above categories, but which can nonetheless lead to the identification of the company, and by extension the people affiliated with the company. For example, in Example 4.1 below, the company Houseplant Inc. was a principal cause of the economic crash in 2030.



### Example 4.1

99. It was, in fact, the statements of Mr. John Doe in his capacity as manager of the Export Division, which the Complainants to relied on when entering into the agreement. His misleading portrayal of the liquidity of Houseplant Inc., and their subsequent bankruptcy led to the total loss of 1.5 billion Euros, and the collapse of the first complainant, which also directly triggered the Third Great Recession of 2030.

....

404. When Houseplant made its infamous televised address to the public, apologising for causing the Third Great Recession of 2030, it was poorly received by governments, and further sanctions were brought on them, including becoming nationalised.

In the above (obviously fictional) example, there are two “well-known” events, which a reader may recognise and connect with the company. The first is events which triggered the Third Great Recession of 2030 (in a way many people today would know Lehman Brothers triggered the Great Recession in 2008). The second is the public apology, which would be recognisable in a way that the BP Oil Spill apology by BP’s CEO would be recognisable to many readers. To approach this, we would redact as below in Example 4.1(a):

### Example 4.1(a)

99. It was, in fact, the statements of Mr. JD in his capacity as manager of the Export Division, which the Complainants to relied on when entering into the agreement. His misleading portrayal of the liquidity of Corp1, and their subsequent bankruptcy led to the total loss of 1.5 billion Euros, and the collapse of the first complainant, which also directly triggered .

....

404. When Corp1 apologising for, it was poorly received by governments, and further sanctions were brought on them, including becoming nationalised.

In this Example 4.1(a), both instances of “the Third Great Recession of 2030” were redacted, although their nationalisation was not. This is because the fact that the company was nationalised may be important to the decision. The fact that Houseplant Inc. is responsible for the Third Great Recession of 2030 would likely be easily traced back to Mr. John Doe. Therefore, the scientific value of leaving the phrase “the Third Great Recession” in paragraph 99 does not outweigh the risk of identifying a natural person – it should be redacted. Additionally, “made its infamous televised address to the public” was redacted as, if it is infamous, it would be recognisable to the reader, and should again be redacted to protect the identity of John Doe.

