



This is an output of K-Peritia CA22101 funded by COST (European Cooperation in Science and Technology) and the European Union

CULTEXP guidelines for GDPR-compliant processing of personal data in legal documents 2026

Noora Arajärvi and Livia Holden ¹

Policy Statement: Governance of Personal Data in Judicial Documents (CULTEXP)

Treatment of personal data under safeguards

CULTEXP operates on the basis that judicial decisions and related legal materials generally constitute personal data within the meaning of Article 4(1) GDPR. Given the narrative and contextual nature of such documents, identifiability may arise not only from direct identifiers but also from the combination of factual elements. Accordingly, the project does not assume that data can be fully anonymised in a stable or irreversible manner.

The processing of these materials is therefore structured as the treatment of personal data under appropriate safeguards, in line with the GDPR's risk-based and accountability-oriented framework.

Pseudonymisation as primary safeguard

The primary safeguard applied is pseudonymisation, understood as the processing of personal data in such a way that individuals cannot be attributed without the use of additional information, which is subject to technical and organisational safeguards. Pseudonymisation is implemented through a combination of measures, including the transformation or controlled removal of identifiers (e.g. through redaction).

Technical implementation: redaction and generalisation

In this context, techniques such as redaction, removal of identifiers, and generalisation are used to reduce identifiability. These measures are understood as technical implementations of pseudonymisation, rather than as independent legal categories. Their effectiveness depends on the broader context of the data and the safeguards applied.

The term “de-identification” is used as an umbrella expression to refer to these technical measures. It does not denote a distinct legal status under the GDPR, which continues to distinguish between personal data, pseudonymised data, and fully anonymised data.

Anonymisation as an exceptional outcome

Anonymisation, understood as the irreversible transformation of data such that individuals are no longer identifiable by any means reasonably likely to be used, is treated as an exceptional outcome rather than the default objective. In practice, particularly in the context of judicial documents, this threshold is difficult to achieve and maintain over time.

As a result, CULTEXP does not rely on categorical claims of anonymisation. Instead, it assumes that most processed materials remain personal data subject to the GDPR.

Risk-based approach and accountability

The level of identifiability is assessed contextually, taking into account the nature of the data, the purposes of processing, and the risks to individuals. Safeguards are calibrated accordingly, with the objective of ensuring that identification is not reasonably likely.

This approach reflects the GDPR's principles of data minimisation, proportionality, and data protection by design and by default (Article 25), as well as the requirement to implement appropriate safeguards in research and archiving contexts (Article 89).

Rather than seeking to eliminate personal data from legal documents, CULTEXP adopts a governance model based on controlled identifiability, in which data remain usable for legal and research purposes while the risks to individuals are effectively mitigated.

¹ This policy is an updated version of the GDPR compliant guidelines for processing personal data in legal documents (2022) <https://culturalexpertise.net/wp-content/uploads/2022/01/gdprcompliantguidelinesforlegaldocuments.pdf>

Table of Contents

Aim and scope of the guidelines.....	2
Treatment of personal data and governance approach.....	2
What is personal data?	3
Legislation and practice on the protection of personal data	5
Pseudonymisation, anonymisation, and de-identification	7
What data require de-identification safeguards?.....	8
What to retain?.....	9
Special categories	10
Challenges and tips	11
Annex 1: Identifying legal persons (as opposed to natural persons) for de-identification.....	12
Annex 2: Practical steps for redacting information related to legal persons	14

Aim and scope of the guidelines

CULTEXP, the main output created by EURO-EXPERT and then further developed by CULTEXP PoC and K-Peritia (CA22101) as 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 processing of personal data in the Member States of the European Union (EU) is regulated by the General Data Protection Regulation 2016/679 (GDPR) and implementing national legislation. The GDPR applies when personal data is processed in connection with the activities of a controller or processor established in the EU, regardless of where the processing actually takes place. Further, the European Union Data Protection Regulation (EUDPR) established by Regulation (EU) 2018/1725 of the European Parliament and the Council of the European Union, sets the rules for how EU institutions, bodies, offices, and agencies process personal data and guarantees individuals' rights regarding the protection of their personal data.

Even a decade after the drafting of the GDPR, interpretation and practices regarding the processing of personal data in legal documents vary greatly, especially between the EU countries, and at times, between different entities and even different courts within one country. The present guidelines for the treatment of personal data contained in the documents archived on CULTEXP have been developed with direct reference to the GDPR, to other regulations and case law.

Treatment of personal data and governance approach

CULTEXP does not assume that data contained in judicial decisions can be fully anonymised in the sense of irreversible de-identification. Instead, the project operates on the basis that such materials continue to constitute personal data within the meaning of Article 4(1) GDPR.

Accordingly, the processing of these materials is structured as the treatment of personal data under appropriate safeguards, in line with the GDPR's risk-based and accountability-oriented framework.

The primary safeguard applied is pseudonymisation, understood as the separation or transformation of identifying elements combined with organisational measures that restrict access to identifying information. Technical measures such as redaction, removal of identifiers, and generalisation are implemented as part of this broader pseudonymisation strategy.

Anonymisation, understood as irreversible de-identification removing data from the scope of the GDPR, is treated as an exceptional outcome rather than the default objective. In most cases, identifiability is reduced but not eliminated,

and data therefore remain subject to GDPR governance, including principles of accountability, proportionality, and data protection by design.

What is personal data?

The legal definition of personal data is set out in Article 4(1) of GDPR. It defines personal data as any information relating to an identified or identifiable natural person. EUDPR applies similar principles to the processing of personal data within the EU institutional system.

Personal data refers to any information that relates to an identified or identifiable natural person. A person is considered identifiable when they can be recognised directly or indirectly through one or more elements linked to their identity. These elements may include obvious identifiers such as a name or identification number, but also less direct identifiers such as location data, professional role, or a combination of factual circumstances that make it possible to single someone out. In practice, personal data therefore covers a wide spectrum of information: from basic identity details to contextual information that, when combined with other data, reveals who a person is.

What matters is not the form the information takes, but its relationship to a specific individual. Personal data can appear in structured databases, written correspondence, emails, audio recordings, photographs, or narrative text. It may consist of objective information, such as a date of birth or employment history, but it can also include opinions, assessments, or descriptions of a person's behaviour. The concept is therefore intentionally broad. Its purpose is to ensure that information connected to an individual is handled with appropriate care and safeguards, regardless of the format in which it appears.

The concept of personal data has also been interpreted broadly in the jurisprudence of the Court of Justice of the European Union. In *Nowak v. Data Protection Commissioner*,² the Court confirmed that personal data is not limited to basic identifying details. Information may constitute personal data whenever it relates to an individual by its content, purpose, or effect.

In addition to the general category of personal data, GDPR recognises certain types of information as particularly sensitive. Under Article 9 of GDPR, special categories of personal data include information revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, as well as genetic data, biometric data used for identification, health data, and data concerning a person's sex life or sexual orientation. The processing of such data is in principle prohibited unless one of the specific legal grounds set out in Article 9(2) applies.

Sensitive data frequently arises in judicial proceedings. Disputes may concern matters such as medical treatment, employment discrimination, family relationships, migration status, or criminal allegations. Judicial files may therefore contain health records, psychological assessments, or other information revealing deeply personal aspects of an individual's life. When such information appears in judgments or procedural documents, its disclosure may have particularly significant consequences for the individuals concerned. For this reason, courts often exercise additional caution when deciding whether and how such data should be included in publicly accessible decisions, for example through more extensive redaction.

The protection of personal data is also closely linked to the protection of vulnerable individuals. Certain categories of persons may face heightened risks if their personal information becomes publicly identifiable in judicial documents. These may include minors, victims of crime, asylum seekers, or individuals involved in family law or protection proceedings. European data-protection frameworks recognise the need for particular safeguards in such contexts. For instance, Recital 75 of GDPR notes that the processing of personal data may pose specific risks to the rights and freedoms of natural persons, especially where sensitive data or information about vulnerable individuals is involved. As a result, courts and judicial bodies often adopt protective measures, such as the use of initials, or restrictions on publication, in order to reduce the risk that individuals can be identified through judicial records while still preserving the transparency and authority of judicial reasoning.

A further category of particularly sensitive information concerns data relating to criminal convictions and offences. Under Article 10 of GDPR, the processing of personal data relating to criminal convictions, criminal offences, or related security measures is subject to specific restrictions. Such data may be processed only under the control of

² Case C-434/16, *Nowak v Data Protection Commissioner*, EU:C:2017:994, 20 December 2017.

official authority or where processing is authorised by statutory provision stemming from EU legislation or Member State law that provides appropriate safeguards for the rights and freedoms of individuals. The rationale for this heightened protection lies in the potentially serious consequences that the disclosure or misuse of criminal information may have for a person's reputation, social standing, and future opportunities.

In the context of judicial documents, information relating to criminal allegations, investigations, or convictions may appear in judgments, indictments, witness statements, or procedural records. Courts must therefore exercise particular caution when publishing such materials. Protective measures such as pseudonymisation, or partial redaction are often used to ensure that the transparency of judicial proceedings does not result in disproportionate exposure of individuals' criminal history or allegations.

In the context of judicial documents – such as judgments, court transcripts and expert reports, the notion of personal data becomes particularly significant. Judicial proceedings inevitably generate large amounts of information about individuals. Court judgments, procedural orders, pleadings, witness statements, and evidence files may all contain references to parties, witnesses, experts, or other persons connected to the case. These references often include identifying details such as names, addresses, professional affiliations, or descriptions of personal circumstances. Even when names are omitted, the factual narrative contained in a judgment can sometimes make it possible to identify individuals indirectly.

Judicial documents may also contain more sensitive forms of personal data. For example, litigation may involve medical information, financial circumstances, employment records, or details of family relationships. Such information may be necessary for resolving the legal dispute, but it nevertheless relates to identifiable individuals and therefore falls within the scope of personal data protection. Courts and judicial bodies must therefore balance two important considerations: the protection of individuals' personal data on the one hand, and the principles of transparency and openness of justice on the other; the treatment of personal data is shaped by this balance. On the one hand, judicial reasoning requires a sufficiently detailed presentation of the facts and the parties involved. Decisions must be understandable, reasoned, and capable of public scrutiny. On the other hand, the publication or dissemination of judicial documents can expose personal data to a wider audience than was originally involved in the proceedings. For this reason, many courts apply measures such as pseudonymisation, or partial redaction when making judgments publicly available. For example, the Court of Justice of the European Union (CJEU) has progressively strengthened its anonymisation practices in recent years – particularly in preliminary ruling proceedings – moving from the use of initials to systematic anonymisation and, more recently, the use of fictitious names, in order to ensure that natural persons cannot be identified when judgments are published.³

Interpretative guidance on the concept of personal data has also been developed by European data protection authorities, including the European Data Protection Board and the European Data Protection Supervisor. A foundational reference remains the Article 29 Working Party's Opinion 4/2007 on the concept of personal data, which explains that information relates to a person when it is linked to them through its content, purpose, or likely effects. This approach is particularly relevant when analysing narrative materials such as judicial decisions, where personal data often appears in contextual or descriptive form rather than as simple identifiers.

In essence, personal data in judicial documents reflects the intersection between individual privacy and the administration of justice. Courts must process personal data as part of their core function of resolving disputes and applying the law. At the same time, they must ensure that such information is handled proportionately and responsibly, particularly when judicial decisions are archived, shared, or published. The concept of personal data therefore serves as a framework that encourages careful consideration of how individuals are represented within legal records and how those records are subsequently used.

A brief clarification is also useful regarding the main actors involved in the processing of personal data. Under the GDPR, the controller is the entity or individual that determines the purposes and means of the processing, while the data processor – when applicable – acts on behalf of the controller and processes data according to its instructions. The data subject is the identified or identifiable natural person to whom the data relates. In the context of CULTEXP,

³ 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, 29 June 2018) <https://curia.europa.eu/site/upload/docs/application/pdf/2018-06/cp180096en.pdf>; Court of Justice of the European Union, 'The Court of Justice introduces fictional names in preliminary ruling cases' (Press Release No 1/23, 9 January 2023) <https://curia.europa.eu/site/upload/docs/application/pdf/2023-01/cp230001en.pdf>; Peter Oliver, 'Anonymity in CJEU cases: the Court changes its approach' (EU Law Analysis Blog, 23 January 2023) <https://eulawanalysis.blogspot.com/2023/01/anonymity-in-cjeu-cases-court-changes.html>.

the project operators or institutions will typically qualify as controllers insofar as they determine how judicial documents are collected, structured, de-identified, and made available, while any technical service providers (for example, hosting or database management) may act as processors. The data subjects, in turn, include all individuals referred to in the underlying materials – parties, witnesses, experts, or third persons mentioned in the factual narrative – whose rights and interests must be taken into account when decisions are made about inclusion, redaction, or dissemination.

Legislation and practice on the protection of personal data

The protection of personal data within the European legal order rests on a normative framework that combines directly applicable Union law, national implementing legislation, and a growing body of interpretative practice. At its core lie GDPR and, for EU institutions and bodies, Regulation (EU) 2018/1725 (EUDPR). These instruments establish not only substantive rights for individuals but also a set of operational principles that must guide any processing of personal data, including in legal and quasi-judicial contexts.

From a doctrinal perspective, the key contribution of the GDPR is the articulation of general principles in Article 5: lawfulness, fairness and transparency; purpose limitation; data minimisation; accuracy; storage limitation; integrity and confidentiality; and accountability. They shape both legislative choices and day-to-day practices. In the context of legal documents, they require a continuous assessment of necessity and proportionality: whether the inclusion of personal data is required for the legal purpose pursued, and whether the same objective could be achieved with lesser means.

Lawfulness of processing, under Article 6 GDPR, is grounded in the performance of a task carried out in the public interest or in the exercise of official authority. This is particularly relevant for courts and public bodies, whose mandate inherently involves the handling of personal data. This legal basis, however, must be read together with the principles of purpose limitation and data minimisation, which constrain the extent to which personal data may be included in publicly accessible documents. The mere fact that information was relevant in the course of proceedings does not automatically justify its full disclosure in a published judgment or database.

Divergent practice across EU Member States demonstrates considerable variation in how these principles are operationalised. Some jurisdictions adopt systematic de-identification of judgments prior to publication, often replacing names with initials or generic descriptors. Others rely on more selective redaction, focusing primarily on sensitive categories of data or vulnerable individuals. There are also differences in the treatment of professional actors: while judges and lawyers are generally named, experts or witnesses may be de-identified depending on the circumstances. This reflects different legal traditions and conceptions of the balance between privacy and the publicity of judicial proceedings.

At EU level, the CJEU has developed its approach and recommendations on de-identification, particularly in preliminary ruling cases involving natural persons, combining guidance to national courts at the referral stage with its own publication practices, including the systematic removal or replacement of identifying details in judgments.⁴ These practices illustrate an important point: data protection in legal documents is not solely a matter of compliance, but also of institutional design. Decisions about what to disclose, and how, are embedded in procedural rules and publication policies.

Supervisory authorities, including the European Data Protection Board and the European Data Protection Supervisor, have further clarified the application of data protection principles to complex contexts. Their guidance consistently emphasises a risk-based approach. This entails assessing not only the nature of the data, but also the likelihood and severity of potential impacts on individuals if the data were disclosed or re-identified. In practice, this shifts the focus from formal categorisation to contextual evaluation: the same piece of information may require different treatment depending on how it is used and combined with other data.

An additional layer of complexity arises from the interaction between data protection law and other legal regimes, notably those governing access to documents, freedom of expression, and the openness of justice. The GDPR itself recognises this interplay, for example in Article 85, which requires Member States to reconcile data protection with

⁴ Court of Justice of the European Union, ‘Recommendations to national courts and tribunals in relation to the initiation of preliminary ruling proceedings’ (2019/C 380/01) [2019] OJ C380/1, <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ%3AC%3A2019%3A380%3AFULL>.

freedom of expression and information. In judicial contexts, this may translate into a structural tension. On the one hand, transparency is essential to the legitimacy and accountability of the legal system. On the other, unrestricted disclosure of personal data may undermine the rights and interests of individuals involved in proceedings.

In practice, the resolution of this tension is rarely absolute. Instead, it is managed through a combination of techniques: pseudonymisation (scope of which is still being defined at the EU judicial level at the time of writing), selective redaction, and controlled access to full records. Importantly, these measures should not be understood as purely technical steps taken at the end of a process. Rather, they reflect a broader compliance logic that should be integrated from the outset: something that the GDPR conceptualises as data protection by design and by default (Article 25).

For CULTEXP, which aggregate and disseminate legal documents across jurisdictions, this has implied an additional responsibility. The act of collecting, structuring, and republishing materials may alter the original context in which the data appeared, thereby increasing the risk of identification or misuse. Consequently, compliance cannot rely solely on the practices of the originating courts or institutions. It requires an independent assessment of whether the data, in its new setting, continues to meet the standards of necessity, proportionality, and fairness set out in the European data-protection framework.

In this context, a further provision of particular relevance is Article 89 GDPR, which introduces a specific regime for the processing of personal data for archiving purposes in the public interest, scientific or historical research purposes, and statistical purposes. Its significance lies not so much in creating a separate legal basis, but in shaping how the general principles of the GDPR are to be interpreted and applied in these contexts.

Substantively, Article 89 recognises that certain forms of data processing – especially those aimed at research, documentation, and long-term preservation – serve important societal functions that may justify a more flexible application of some data protection constraints. This is particularly pertinent for databases such as CULTEXP, which operate at the intersection of legal documentation, academic research, and public access. However, Article 89(1) requires that such processing be subject to “appropriate safeguards”, with a specific emphasis on technical and organisational measures that ensure respect for the principle of data minimisation. In practice, this often translates into pseudonymisation, implemented through measures such as redaction, generalisation, and access restrictions.

Article 89 can operate as a facilitating provision. It allows data controllers to retain and reuse data for research or archiving beyond the original purpose for which the data was collected, provided that the safeguards are in place and that the further processing is not incompatible with the initial purposes (Article 5(1)(b)). This is a critical point for legal databases: judicial decisions and expert reports are typically produced for the resolution of a dispute, yet their subsequent use in research infrastructures or comparative legal analysis can be justified under Article 89, provided that the risks to individuals are adequately mitigated.

Interpretatively, Article 89 has been read by supervisory authorities and in the literature as embedding a form of proportionality test specific to research and archiving contexts.⁵ It does not displace the core GDPR principles but modulates their application. For example, certain data subject rights – such as the right of access, rectification, restriction of processing, and objection – may be derogated from under Article 89(2), where their exercise would seriously impair or render impossible the achievement of the research or archiving purposes, and where such derogations are provided for in Union or Member State law. These limitations must remain exceptional and justified and cannot be invoked in a blanket manner, and they do not affect the overarching obligation to implement safeguards.

The practical consequence is that Article 89 introduces a structured flexibility rather than an exemption. It requires controllers to actively demonstrate that the chosen balance between data utility and data protection is justified. In the context of legal documents, this often involves a nuanced assessment: full anonymisation (where achievable) may reduce risks but also diminish the analytical value of the material, particularly where contextual or relational elements

⁵ General Data Protection Regulation Art. 89(1)-(2) and recital 156; European Data Protection Board, *Legal study on the appropriate safeguards under Article 89(1) GDPR for the processing of personal data for scientific research* (2021) https://www.edpb.europa.eu/system/files/2022-01/legalstudy_on_the_appropriate_safeguards_89.1.pdf; see also Christian Wiese Svanberg, ‘Article 89 GDPR. Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes’ in Christopher Kuner et al. (eds.), *General Data Protection Regulation (GDPR): Commentary* (OUP 2020).

are essential. On the other hand, retaining too much identifiable information may expose individuals to unnecessary risks, especially when data is made openly and publicly searchable.

The potential of Article 89, therefore, lies in enabling responsible data reuse.⁶ Properly applied, it supports the development of comprehensive legal and research databases while maintaining a high level of protection for individuals. It encourages a move away from binary approaches to disclosure towards more nuanced, risk-based solutions combining technical and organisational safeguards, such as partial de-identification and controlled access. For CULTEXP, this provision provides a legal and conceptual foundation for its core activity: making legal materials accessible for research and comparative purposes, while embedding safeguards that reflect both the sensitivity of judicial data and the evolving expectations of data protection in the digital environment.

In sum, legislation provides the normative architecture, but the evolving practice gives it operational meaning. Effective protection of personal data in legal documents depends on the ability to translate abstract principles into context-sensitive decisions that respect both individual rights and the functional requirements of the legal system.

Pseudonymisation, anonymisation, and de-identification

The concept of personal data under the GDPR is broad and includes both direct identifiers (such as names or identification numbers) and indirect identifiers (such as location, profession, or contextual elements that may enable identification). In judicial documents, identifiability often arises from the combination of narrative details rather than from isolated data points.

In these guidelines, ‘de-identification’ is used as an umbrella term referring to technical measures (such as redaction and generalisation) implemented within a broader pseudonymisation framework. It does not denote a distinct legal category under the GDPR.

Within this framework, it is important to distinguish between anonymisation and pseudonymisation measures, implemented through techniques such as redaction, generalisation, or removal of identifiers.

Anonymisation refers to the irreversible transformation of data such that individuals are no longer identifiable by any means reasonably likely to be used. Where this threshold is met, the data fall outside the scope of the GDPR. However, in practice, particularly in the context of judicial decisions and narrative materials, this threshold is difficult to achieve and maintain over time.

Pseudonymisation, by contrast, consists of processing personal data in such a way that they cannot be attributed to a specific individual without the use of additional information, which is kept separately and subject to safeguards. Under the GDPR, pseudonymised data remain personal data but benefit from reduced risk when appropriate technical and organisational measures are applied.

In CULTEXP, pseudonymisation constitutes the primary governance approach, implemented through pseudonymisation measures such as redaction, generalisation, or removal of identifiers. Identifiability is reduced through a combination of measures, including the removal or transformation of identifiers, controlled access, and contextual assessment of risk.

Redaction is understood as a technical operation consisting in the removal or masking of specific elements of a dataset, such as names, dates, or locations. While redaction plays a central role in reducing identifiability, it does not in itself determine whether data remain personal data under pseudonymisation safeguards or reach the threshold of anonymisation. Its effectiveness depends on the broader context and the safeguards applied.

As a result, CULTEXP does not rely on categorical claims of anonymisation. Instead, it adopts a risk-based approach in which personal data are processed under pseudonymisation safeguards, and identifiability is mitigated through layered technical and organisational measures.

⁶ Rossana Ducato, ‘Data protection, scientific research, and the role of information’ (2020) 37 *Computer Law & Security Review* 105412.

What data require de-identification safeguards?

As clarified above, ‘de-identification’ refers to technical measures implemented within a broader pseudonymisation framework. The following categories of data require risk-based mitigation of identifiability (i.e. de-identification measures), typically through pseudonymisation measures such as redaction, generalisation, or removal. The objective is not necessarily to eliminate all possibility of identification, but to ensure that identification is not reasonably likely, taking into account the context of publication and use.

1. Names of **all** living individuals (data subjects) 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 de-identification. 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^, 20 ottobre 1999, Bajrami" (or with the name of judge).

 "Cass. Sez. 6^, 20 ottobre 1999, B[removed]"

What to retain?

The following data should normally be retained, unless they create a risk of identification requiring mitigation:

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, 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 (unless revealing personal data).
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 de-identifying).
 - 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.

Special categories

As a general principle, the GDPR applies to the processing of personal data within its material and territorial scope. At the same time, it leaves a degree of discretion to Member States in its implementation, which has resulted in significant variation in national legislation and practice, particularly in the context of judicial documents. Questions such as whether to de-identify judges do not have a uniform answer, as they require balancing data protection with principles such as transparency and the openness of justice.

While relevant guidance and case law exist, there is no harmonised or binding EU-level rule that resolves these issues in a comprehensive manner. In practice, this leads to context-specific approaches, where the treatment of personal data depends on the function of the individual concerned and the purpose of publication. In many cases involving judicial activity, the processing is more appropriately grounded in Article 6(1)(e) GDPR, as it forms part of a task carried out in the public interest. This requires a proportionality assessment, taking into account the role of the individuals concerned. With regard to court staff, and in particular judges acting in their official capacity, the interest in transparency and accountability will often weigh strongly in favour of publication, although this must be assessed on a case-by-case basis. The publication of judges' names contributes to the openness of justice and democratic debate, and may also fall within the scope of Article 85 GDPR, which requires a reconciliation between data protection and freedom of expression and information.

The principle of transparency in the judicial process, and the broader principle of open justice reflected in Article 47 of the Charter of Fundamental Rights of the European Union, support the publication of the names of judges as part of ensuring accountability, and reinforcing the 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.⁷ Further, it has been observed that “in most countries the names of professionals involved [in a court case] are not rendered anonymous [...]”.⁸ Hence, as a general rule, the people involved in a court case in their professional capacity are not

⁷ 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>.

⁸ 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/JILawInfoSci/2016/3.html#fn66>.

de-identified in CULTEXP; for example judges, clerks, lawyers, bailiffs, court experts, interpreters and custodians.⁹ Exceptions to this may include specific country practices – the respective Data Protection Authorities of different EU 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).¹⁰

Given the variation in national practices and the evolving nature of guidance in this area, the project adopts a cautious and adaptive approach. It follows applicable national legislation and regional practices, including where these require the redaction of personal data relating to judges and court staff. At the same time, it continuously monitors developments in the interpretation and implementation of the GDPR, with a view to adjusting its practices as further guidance and standards emerge.

Challenges and tips

Removing direct identifiers is not enough to ensure that identification of the data subject is no longer possible!

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 identification set size is three and a half billion, but if he is described as ‘a middle-aged Dutchman with a beard’ the identification 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 identification 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 the level of identifiability and the safeguards applied (pseudonymisation, redaction, generalisation), and whether full anonymisation is exceptionally achieved.
- 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 are retained without mitigation, explain why you cannot de-identify 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.

⁹ Marc 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.boeccli.eu/uploads/deliverables/Deliverable%20WS0-D1.pdf>, p. 24.

¹⁰ Tristan Allard, Louis Béziaud and Sébastien Gambs, ‘Online publication of court records: circumventing the privacy-transparency trade-off’, 3 Jul 2020, <https://arxiv.org/abs/2007.01688>.

- Assess whether the remaining data combined with additional information might permit the identification of an individual.
- Prepare a short statement disclosing your de-identification 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 (as opposed to natural persons) for de-identification

1. Introduction

There are several reasons for 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 CULTEXP, and (2) to remove any data which could reveal information about unrelated parties.

The need to 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 redact the natural person's name and relationship with the legal person, or (2) to 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 de-identify 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 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).

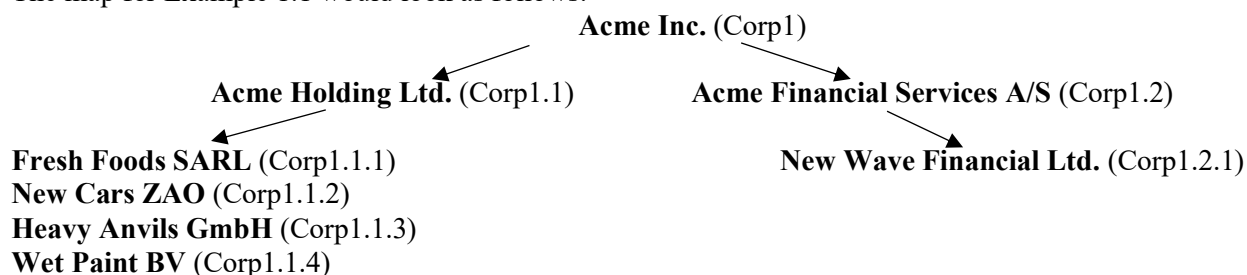
1. Mapping out the corporate structure

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.

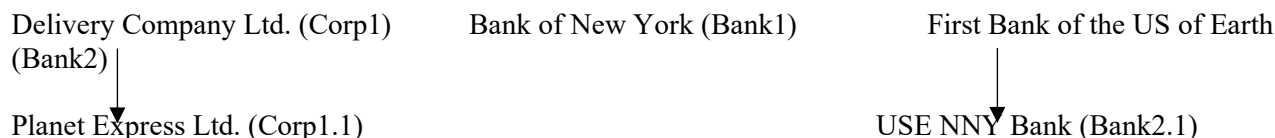
2. Outside corporations and groups

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:



Good News Ltd. (Corp2)

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’ 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.