Guidance Document: *Data Suppression and Obfuscation*

Parent Document

State of Indiana Policy: Information Privacy.

General Guidance

Data deidentification is a spectrum. Many regulatory frameworks protect sensitive information similarly. On one end, we have directly-identifiable data; this is source data that includes details like first and last name, data of birth, and perhaps social security number. With it, we include indirectly-identifiable data, which generally includes information that describes, locates, or indexes anything about an individual, or that affords a basis for inferring personal characteristics about an individual. On the other end, we have truly anonymized data; this is derived from source data and is often aggregated and the groups suppressed to ensure that small counts cannot be used to reidentify an individual in an aggregated grouping. The former is heavily protected under many regulatory frameworks while the latter may be released openly.

In the middle, we have shades of gray. Classifications of data that are either likely or unlikely to be identifiable, based on factors like the scope of data, context of its use, and other data available to the user that is linked or linkable with the data in question. It is this middle ground where the heavy lifting takes place. This is the data that, due to degrees of deidentification, is more available and valuable to researchers, but explicit regulatory guidance—a hard 'yes' or 'no' with respect to releasability—is often lacking.

In what has been described as "a dogmatic compliant approach," legal counsel uncomfortable with data protection law and privacy-enhancing technologies will decree that this middle ground, like directly-identifiable data, is off limits to researchers and policymakers. This approach is overly-restrictive, enabling the release only of data that is truly anonymized or that which is subject to an explicit regulatory exception—one that allows data to be shared with a specific individual or entity for a specific purpose. This approach renders valuable data useless in a data-driven organization and is a hallmark of old thinking.

Indiana State Government is embracing the efficient and ethical use of data to drive decisions and improve outcomes for Hoosiers. To that end, where legal counsel embraces the shades of gray in data protection law and leverages privacy-enhancing technologies, we find data-driven organizations whose data culture thrives, while respecting the privacy rights of individuals and complying with relevant regulatory frameworks. This is the goldilocks zone of data deidentification and an enabler of a robust data culture.

The Information Commissioner's Office of the United Kingdom has published a draft anonymization decision tree, which clarifies the lines between these deidentification lanes. See Figure 1, below.

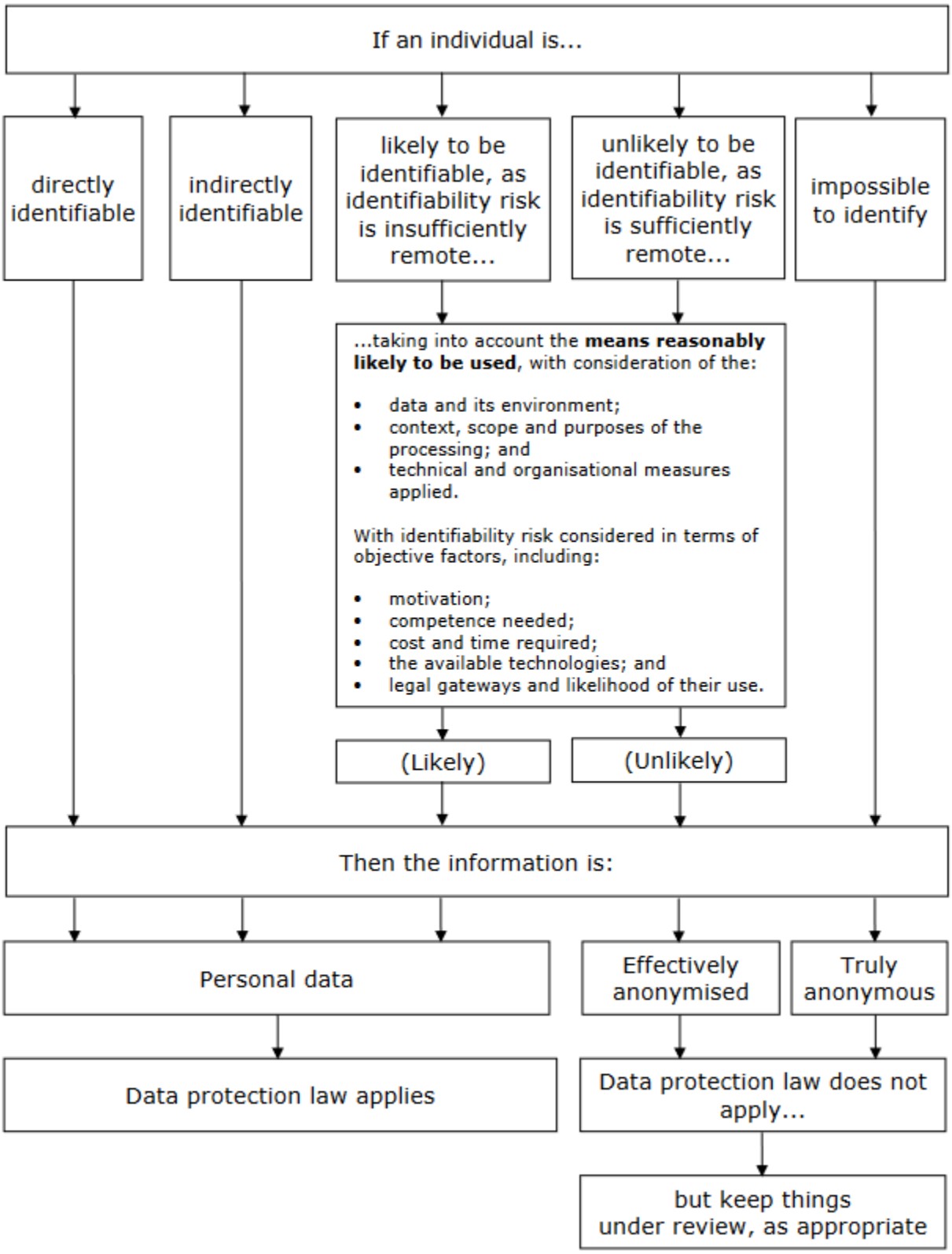

Figure 1. Info Comm'rs Office, Anonymisation Draft, 2021, licensed under the Open Govt. License.

The Indiana Office of Chief Data Officer maintains specialized legal and technical competency related to data sharing and deidentification and is available to assist agencies with the data release process through the Indiana Management Performance Hub Data Review Team and OCDO Privacy Board.

Public Release Guidance

In the United States, data privacy standards are sectoral, meaning that they reside within specific domains, like health and human services, education, or drivers' privacy. As a result, no single obfuscation standard applies to all Personal Information. This Guidance Document offers the following general recommendations to agencies wishing to obfuscate Personal Information prior to its public release:

1. In the case of aggregate information, suppress the information so that groups of "n" counts fewer than ten (10) are obfuscated and apply secondary suppression as needed to ensure that suppressed cells fewer than ten (10) may not be recalculated through subtraction using the remaining cells. Finally, consult the General Guidance above as well as the law or regulation which governs the information proposed for release and ensure that the data product either: 1) no longer constitutes a protected class of information; or 2) qualifies for a disclosure exception based on its content, the ultimate receiver, and the receiver's proposed use of the information.
2. In the case of row-level information, consult the General Guidance above as well as the law or regulation which governs the information proposed for release. Remove or obfuscate data elements in a manner so as to ensure that the data product either: 1) no longer constitutes a protected class of information; or 2) qualifies for a disclosure exception based on its content, the ultimate receiver, and the receiver's proposed use of the information.

HIPAA Guidance

The Health Insurance Portability and Accountability Act of 1996, or HIPAA, furthers this deidentification approach. Pursuant to HIPAA, one of two processes must be completed to enable the public release of protected health information, or PHI. The first process is known as "expert determination." Expert Determination requires the following:

> A person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable: (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and (ii) Documents the methods and results of the analysis that justify such determination.

45 CFR 164.514(b)(1).

The second process is referred to in the HIPAA Privacy Rule as "safe harbor." Safe Harbor requires the removal of 18 enumerated data elements, rendering the data no longer PHI and thus eligible for public disclosure.

There are benefits and risks associated with each of these approaches. For instance, while Safe Harbor provides a clear path to deidentification, the resulting data product may be less useable for a particular use case due to the high degree of obfuscation. While Expert Determination may enable the release of a more complete data product, the level of effort to meet the Expert Determination requirements is significant. The OCDO, in partnership with the Indiana Family and Social Services Administration, has developed a HIPAA Expert Determination methodology for Indiana State Government, which can guide these types of initiatives. For more reading on this subject, consult the *OCDO Standard: HIPAA Deidentification Methodology*, which is available on the Indiana Privacy Program webpage at on.IN.gov/privacy.

Closing

In all of these cases, the Indiana Office of Chief Data Officer maintains specialized legal and technical competency related to data sharing and deidentification and is available to assist agencies with the data release process through the Indiana Management Performance Hub Data Review Team and OCDO Privacy Board. Please contact us for more information.