AIDA GDPR policy

Introduction

Analytic Imaging Diagnostics Arena (AIDA) is a strategic initiative within Medtech4Health, supported by Vinnova. This policy describes how the AIDA community interprets some key aspects of GDPR and national Swedish legislation in relation to typical research and innovation scenarios in AIDA. The policy has been developed jointly by the AIDA community, has been reviewed by the legal department at Linköping University and has been approved by the AIDA Steering Group. The AIDA community has concluded that the provisions in this policy reflect an integrity-preserving, legal, and ethical handling of research data.

Whenever a situation or GDPR aspect is not covered by this policy, no assumptions can be made about how the AIDA community would define recommended handling. This policy does not mean that AIDA takes on any responsibilities, this remains with the organization where an AIDA researcher is employed.

Lawfulness

The lawfulness of processing personal data for research is established by GDPR Article 6.1e:

> *"[Processing shall be lawful if] processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller."*

Thus, there is a general permission to conduct research with personal data within AIDA, if it is done in compliance with the other provisions within GDPR. Note that the above means that GDPR does not necessarily require consent. (Consent is another way to make processing lawful, as defined in Article 6.1a.)

Currently, the Swedish legislation is not finalized regarding whether research by non-universities such as private companies can be covered by public interest lawfulness. The proposal from the legislative process (Forskningsdatautredningen) and the position of Vetenskapsrådet is that research is of public interest also when conducted by private companies. Awaiting finalized legislation, the AIDA policy concurs with this view.

Definition of anonymous data

GDPR only applies to *personal data*. If the data is anonymous, GDPR does not come into play. Personal data is defined (Article 4.1) as information making it possible to identify a person directly or indirectly, for instance through use of additional information from other sources. To determine if it is possible to identify a person one should consider all means *reasonably likely* to be used (Recital 26). How to determine reasonable likeliness is further explained as follows (Recital 26):

> *"…account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments."*

A fundamental question for AIDA is whether an image extracted from a clinical source can be considered as anonymous, when the original remains in the clinical database in which it, naturally, is connected to a patient identity. While it can be concluded that an image copy is in itself a unique key linking back to the clinical original, the anonymity status can still be motivated by GDPR's likeliness criterion. This AIDA policy concludes that **images extracted from clinical databases can be considered anonymous** if handled thoughtfully.

With respect to anonymization, the AIDA interpretation of the GDPR's "*reasonably likely"* term is as follows:

> *Data is anonymous when the de-identification procedure results in that there is no reasonably practical way for anyone, not even a care provider employee or IT support staff of the care provider, to reconnect the data (the image and/or the meta data) to the clinical record.*

This means that to ensure anonymity, meta data that in a wide sense can be considered to contribute to identifiability must be removed. Obviously, this includes direct patient identifiers, but also other information that could be used for direct searches in a clinical database finding a narrow group of patients. One guidance can be found in the HIPAA regulations in the US where 18 identifier types to be removed are listed. It is not, however, sufficient to consider predefined lists. A thorough analysis of all image and meta data must be done to spot potential identifiers. For example, examination date in combination with modality or lab information may narrow a patient search to just a few hits. Likewise, it can be an issue if a disease is named and it is very rare, especially in combination with other information such as the age of the patient and the name of the hospital. Volumetric imaging may be possible to reconstruct into identifiable faces, and there may be other unique characteristics on the body such as birthmarks or tattoos represented in the image data.

The above reasoning entails that the following scenario is **not** considered reasonably likely:

> *A person finds out which care provider the image is from, unlawfully achieves unrestricted and long-time access to the clinical database system, uncovers the detailed technical architecture of the clinical database system, develops an image-matching search procedure specific for the architecture, deploys the procedure in an exhaustive search across millions of images, and thereby identifies the person.*

Apart from not being reasonably likely, it can be argued that the above scenario also makes the question of anonymous research data irrelevant. If a person would have such access as described above, then all patient data in the system would be accessible anyway, and no additional integrity risk arises from having a research copy of some of the clinical data.

To avoid any doubt, this AIDA policy also concludes that anonymization in itself (actions on personal data carried out for the purpose of anonymization) in order to use the data for research, is to be considered lawful. Moreover, this policy concludes that employing anonymization, when the research agenda allows, is a suitable way to follow the data minimization principle in GDPR.

Processing safeguards

In GPDR Article 89.1, it is stated that appropriate protective technical and organizational measures are to be taken to safeguard against improper use of personal data. One such measure applied to all data hosted in the AIDA environment is that data is stored in a protected cloud infrastructure (at the time of writing supplied by Microsoft). An organization needs to have a formal AIDA affiliation to be allowed to access data hosted in the AIDA environment, and the AIDA system will keep an audit log that registers which parties that have (actively) accessed which data. This safeguard is applied also to anonymized data.

An AIDA affiliate accessing anonymized AIDA environment data from other parties is required to sign an agreement not to spread data further and to delete their copies of the data if mandated by AIDA management. This safeguard is needed in case a judgement that data is anonymous is overruled at a later time.