|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-J-054 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | Plenary | E-meeting, 30 September – 2 October 2020 | |
| **DOCUMENT** | | | | |
| **Source:** | | Editors | | |
| **Title:** | | Updated DEL5.6: Data sharing practices | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Ferath Kherif Centre Hospitalier Universitaire Vaudois Switzerland | | Email: [Ferath.Kherif@chuv.ch](mailto:Ferath.Kherif@chuv.ch) |
| **Contact:** | | Banusri Velpandian ICMR India | | Email: [banusrir@gmail.com](mailto:banusrir@gmail.com) |
| **Contact:** | | WHO Data Team | | Email: [pujaris@who.int](mailto:pujaris@who.int) |

|  |  |
| --- | --- |
| **Abstract:** | This initial draft describes the objectives and proposes an initial outline of the planned deliverable “Data Sharing Practices” to help seed future content. |

**Initial outline**

# Introduction

This deliverable aims to provide an overview of the existing best practices for data sharing of health-related data. The scope of this document includes a description of all the necessary steps and requirement to enable secure data sharing. The document specifies the role of the data providers, data processors and the data receivers. The document outlines established data sharing methods and novel methods based on distributed and federated environments for privacy preserving AI/ML models.

# Why data sharing and data sharing principle

Sharing health-related data at large scale is needed to accelerate the development and the use of AI solutions for health.

# Data sharing principles

The Data sharing principles described here relate to the overall data management of including availability, usability, integrity and security of the data shared between organizations.. Examples of current principles includes FAIR Principles and DATS standard to describe / search datasets

# The specificity of sharing health-related data

In the context of AI solution for healthcare application these principles need to be extended to issues related to privacy, standardization and interoperability.

# Examples of data sharing scenarios

# Research data

Data is accessible in public repositories.

# Clinical and hospital data

Data is stored in local private databases (e.g. hospitals)

# Roles and responsibilities

# Data providers

Natural or legal person who collects data as part of their working duties. Data producers operate under the control of Data Controller

# Data protection officer

Person responsible for overseeing data protection strategy and compliance with regulation (e.g. GDPR)

# Data controllers

Natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law (Article 4(7)).

# Data processors

Natural or legal person, public authority, agency or any other body, which processes personal data on behalf of the controller.

# Data receivers

Natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not.

# Rules and regulation for patients consent

Data collected from human beings must have been collected according to the ethical principles governing research in the EU. Where the data were collected as part of the HBP work, compliance with the ethical principles will have been checked during the Ethics Review.

Evidence must be provided that:

1. The data subject consented to the procedure undertaken to collect the data
2. The data subject consented to the use of the data for the research purposes that it is to be used for.
3. Where no consent for data sharing is available, the re-use of the data must be legal according to EU data protection legislation.

These principles apply to both experimental data collected from volunteers and to medical and patient data. They are equally valid for identifiable human data and anonymized and / or aggregated data. Data that is not identifiable may not fall under data protection regulation, but this will need to

# Data anonymization and de-identification

Removal of any information relating to an identified or identifiable natural or legal entity, including institutions and animals. An identifiable entity is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to his or her physical, physiological, genetic, mental, economic, cultural, or social identity.

# Data minimization

One of the most important principle is the data minimization where collection of personal data must be limited to data that is relevant, adequate, and absolutely necessary for carrying out the purpose for which the data is processed.

# Data confidentiality, Data security and privacy

The aim of the Data protection is defined as keeping information secure so that it is only used for the defined, purposes and it is stored, managed and used in a way that ensures the privacy of the data subjects involved, and assures integrity of the information so that the information are correct.

# Data descriptions, Metadata registry and Data harmonisation

* + Use/adapt exiting standards (e.g. FDA,CDISC)
    - Description of a cohort e.g. DCAT vocabulary
    - Description of datasets e.g. STDM Study Data Tabulation Model
  + Explore and test exiting software implementation for master data management
    - Mobi [https://mobi.inovexcorp.com](https://mobi.inovexcorp.com/)
    - Apache Atlas [https://atlas.apache.org](https://atlas.apache.org/)
    - Magda [https://magda.io](https://magda.io/)
    - Egiria [https://egeria.odpi.org](https://egeria.odpi.org/)

# Methods for data sharing

# Overview of data transfer methods

# Central and public databases

# Distributed and federated query

# Distributed and federated machine learning

# Data governance in a centralized environment and governance in a federated environment

# Legal documents

# Data protection impact assessments (DPIA)

DPIA is a process designed to describe the processing, assess the necessity and proportionality of a processing, and to help manage the risks to the rights and freedoms of natural persons resulting from the processing of personal data (by assessing them and determining the measures to address them). Privacy impact assessment (PIA) under the GDPR is a requirement. Data Protection impact assessments must be carried out by the data controller or the processor acting on the controller’s behalf. The DPIA report identify privacy risks and how to mitigate them.

* 1. General Data Protection Regulation (GDPR)

The General Data Protection Regulation (EU) 2016/679 (GDPR) is a regulation in EU law on data protection and privacy in the European Union (EU) and the European Economic Area (EEA).

# Data sharing agreement

Data sharing Agreements determine how and for which purposes data are held in the relevant services. The Agreement should include the following information:

* + Purpose and intended use of data sharing
  + Period of agreement
  + Description of data
  + Data update
  + Responsibilities of data providers and data receivers
  + Results and dissemination of results including IA models

# Risks, costs and benefits

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_