|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-L-044 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | Plenary | E-meeting, 19-21 May 2021 | |
| **DOCUMENT** | | | | |
| **Source:** | | Editors DEL5.6 | | |
| **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. This deliverable aims to provide an overview of the existing best practices for data sharing of health-related data, including the requirement to enable secure data sharing and issues related to data governance. The documents described established solutions and novel approaches based on distributed and federated environments. |

# Introduction

AI solution developers for healthcare understand what an important role sharing data plays in their success. In addition to patients, healthcare organizations, government agencies realize the value of sharing data when considering the beneficial outcomes.

This deliverable provides guidance for existing industry best practices for the sharing of health-related data. It outlines the roles of each party with respect to the data provider, processor, and receiver while exploring traditional and novel approaches leveraging distributed and federated methods for developing privacy-preserving AI/ML models.

# Data sharing

## Principles

Sharing medical data at scale is necessary to improve the development and adoption of artificial intelligence solutions for healthcare and to make these solutions more robust. Healthcare, more than any other industry, must negotiate the balance between potential harm from sharing critical data and the potential benefit of improving care. In consideration of this balance and ethical and security issues, several government agencies have issued guidelines that can be summarized under five data sharing principles.

* **Sharing data for appropriate and authorised purposes**. *Why the data is being used*
* **Sharing data only with authorised users.** *Who is using the data*
* **Using data in a safe and secure environment.**
* **Applying appropriate protections to the data.** *Where the data is being used*
* **Ensuring public outputs from data sharing projects do not identify the people or organisations in the data.**

## Roles and responsibilities in handeling personal data

There are several roles and responsibilities associated with the various types of data handling during the life cycle of data sharing identified by GDPR. The data controller is the individual responsible for determining the purposes and means of processing personal data. Data providers are natural or legal persons who collect data as part of their working duties. It is the responsibility of the Data Protection Officer to oversee the organization's data protection strategy and to ensure compliance with regulations such as GDPR. A Data Processor is someone, a public authority, an agency or any other entity that processes personal data on behalf of the controller. Finally a Data receiver is a natural or legal person, public authority, agency, or other body to which the personal data will be disclosed, whether a third party or not.

## Rules and regulation for patients consent

In order to collect data from humans, they must have been collected in a manner that meets the ethical principles governing research in the EU. It must be demonstrated that the patient consented to the collection of data. Data collected from patients is only used for the purpose for which they were collected. In certain instances, the re-use of non-personal data could be allowed under EU data protection legislation, including the recent Data Protection Act in the EU, in which no consent is required to share the data.

## Data anonymization and de-identification

Remove all data relating to identifiable natural or legal entities, including institutions and animals. An identifiable entity is one who can be identified directly or indirectly, in particular by reference to one or more factors specific to their physical, physiological, genetic, mental, economic, cultural, or social identity.

## Data minimization

In order to protect data privacy, data minimization means that data collection must be limited to information that is relevant, adequate, and absolutely necessary for the purpose for which it 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.

## Protection and Privacy assessments.

Data protection impact assessments (DPIA) is a tool used to describe a processing, assess its necessity and proportionality, and help manage risks to the rights of natural persons resulting from processing personal data. The DPIA report identify privacy risks and how to mitigate them.

## 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

# Data Sourcing

## Principle

Data sourcing is aligned with the goal of the AI4Health focus group, an initiative of the International Telecommunications Union (ITU) of the United Nations which is to enable technical and regulatory tools for deployment, assessment, and clinical validation of healthcare AI for medicine.

## Data Access, Quality and Curation

DAQCORD Indicators: descriptive system for planning and reporting observational studies

[**https://www.daqcord.org/daqcord-questions**](https://www.daqcord.org/daqcord-questions)

## Data descriptions, Metadata registry and Data harmonisation

## Methods for data sharing: Central and public databases

## Methods for data sharing: Distributed and federated Access

## Data catalog

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

**Step and Timescale**

|  |  |  |  |
| --- | --- | --- | --- |
| Sr No | Tasks name | Description | Status |
| 1 | Data ingestion pipeline- capturing metadata | The best practices for automating the extraction of metadata from headers or during pre-processing should be defined and presented. | Ongoing  End Date (v1): 28.04.2021 |
| 2 | Metadata capture form- develop draft (DS and Clinical input for questionnaire) | Develop a business process to capture the metadata from domain experts (clinicians and data scientists). Collaborate with experts (Topic groups leaders) to establish, improve, and test the metadata workflow. | Ongoing  End Date (v1): 28.04.2021 |
| 3 | Metadata capture form- DI pipeline | We will demonstrate what is required to complete step 1 (Data Ingestion) pipeline- capturing metadata) and apply this approach to an exemplary dataset. | Ongoing  End Date (v1): 05.05.2021 |
| 4 | Identify public datasets-10 (include imaging data, vitals, etc) | Select ten examples that should reflect the diversity of sizes, application domains, and other aspects of public datasets. | Ongoing  End Date (v1): 28.04.2021 |
| 5 | Structure Metadata for public datasets | We will demonstrate what is required to complete step 2-3 and apply this approach to the selected datasets. | Ongoing  End Date (v1): 15.05.2021 |
| 6 | Ingest Data with associated metadata | TBD |  |
| 7 | Test ingestion of some metadata as part of data ingestion | TBD |  |
| 8 | Fill associated metadata form | TBD |  |
| 9 | Identify research papers with related AI algorithm | TBD |  |
| 10 | Do a quick relevant datasearch using metadata tags | TBD |  |
| 11 | Customer Feedback- Clinical Evaluation and Data Science Team | TBD |  |

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