|  |  |  |
| --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION**TELECOMMUNICATIONSTANDARDIZATION SECTOR**STUDY PERIOD 2017-2020 | FG-AI4H-G-205-A03 |
| **ITU-T Focus Group on AI for Health** |
| **Original: English** |
| **WG(s):** | Plenary | New Delhi, 13-15 November 2019 |
| **DOCUMENT** |
| **Source:** | Editor |
| **Title:** | DEL05C: Data annotation specification |
| **Purpose:** | Discussion |
| **Contact:** | Shan Xu, CAICT, China | Email: xushan@caict.ac.cn  |
| **Contact:** | Harpreet Singh, ICMR, India | Email: hsingh@bmi.icmr.org.in  |
| **Contact:** | Secretariat of the Focus Group | Email: tsbfgai4h@itu.int  |

|  |  |
| --- | --- |
| **Abstract:** | This initial draft describes the objectives and proposes an initial outline of the planned deliverable “Data Annotation Specification” to help seed future content. The objective of the deliverable is to provide a brief introduction of the definition, importance, general framework, procedure, requirements of data annotation in artificial intelligence for health. |

**Initial outline**

# Introduction

* What is data annotation?
* Type of biomedical research data
* Working definition of data annotation
* Importance in data annotations particularly in medical research
* Procedure/SOPs for data annotations
* Examples

<TBD>

# Scope

The data annotation specification gives a brief introduction of the definition, importance, general framework, procedure, requirements of data annotation in artificial intelligence for health.

<TBD>

# Reference

The following ITU-T Recommendations and other references contain provisions, which, through reference in this text, constitute provisions of this document. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; readers of this document are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below.

A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this document does not give it, as a stand-alone document, the status of a Recommendation.

* Available specifications such as ICF, ICD etc.

<TBD>

# Definitions

* Defining data annotation specifications
	+ Controlled vocabulary
	+ Metadata
	+ Examples

# Terms defined elsewhere

This document uses the following terms defined elsewhere:

<TBD>

# Terms defined in this Recommendation

This document defines the following terms:

<TBD>

# Abbreviations and acronyms

This document uses the following abbreviations and acronyms:

<TBD>

# Background

<TBD>

* Present status and future directions
	+ Semantic ontologies and integration
	+ Individual researchers vs consortium
	+ Professional paid services

# General requirements of data annotation

# Types of data annotations

* Human readable
* Machine readable
* Hybrid

# Data privacy

<TBD>

# Quality requirements of data sets

<TBD>

# Qualification requirements of labelling doctors

<TBD>

# Requirements of data distribution

In order to avoid bias in data distribution, it is suggested that the labelling of medical image is carried out by grouping and crossing. Figure 8.1 illustrates the recommended data distribution workflow.



Fig 8.1 Workflow of Data Distribution

<TBD>

# Requirements of data annotation tool

<TBD>

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