|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU Logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION STANDARDIZATION SECTOR**  STUDY PERIOD 2017-2020 | | FG-AI4H-G-205 | |
| **ITU-T Focus Group on AI for Health** | |
| **Original: English** | |
| **WG(s):** | | Plenary | New Delhi, 13-15 November 2019 | |
| **DOCUMENT** | | | | |
| **Source:** | | Editor | | |
| **Title:** | | DEL05: Data Specification | | |
| **Purpose:** | | Discussion | | |
| **Contact:** | | Marc Lecoultre ML|LAB.AI Switzerland | | Email: [ml@mllab.ai](mailto:ml@mllab.ai) |
| **Contact:** | | Secretariat of the Focus Group | | Email: [tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int) |

|  |  |
| --- | --- |
| **Abstract:** | The present document proposes an outline for the future deliverable "Data Specification". Background: The ITU/WHO Focus Group on Artificial Intelligence for Health (AI4H) has proposed a list of deliverables at meeting "G" in New Delhi in November 2019, including this "Data Specification", which combines a set of four deliverables as umbrella. |

Call for Participation

If you are interested in contributing to the *AI4H Data Specification*, please contact the editor of this document (Marc Lecoultre, [ml@mllab.ai](mailto:ml@mllab.ai)) and the Secretariat of the Focus Group ([tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int)) with "ITU/WHO Deliverable - Data Specification" as e-mail-subject, briefly introduce yourself, describe your relevant expertise, and explain your interest.

Table of Contents

[Objectives 3](#_Toc27141523)

[1 Data requirements 4](#_Toc27141524)

[1.1 Objectives 4](#_Toc27141525)

[1.2 … 4](#_Toc27141526)

[1.3 … 4](#_Toc27141527)

[1.4 Bibliography 4](#_Toc27141528)

[2 Data acquisition 5](#_Toc27141529)

[2.1 Objectives 5](#_Toc27141530)

[2.2 … 5](#_Toc27141531)

[2.3 … 5](#_Toc27141532)

[2.4 Bibliography 5](#_Toc27141533)

[3 Data annotation specification 6](#_Toc27141534)

[3.1 Objectives 6](#_Toc27141535)

[3.2 … 6](#_Toc27141536)

[3.3 … 6](#_Toc27141537)

[3.4 Bibliography 6](#_Toc27141538)

[4 Training and test data specification 7](#_Toc27141539)

[4.1 Objectives 7](#_Toc27141540)

[4.2 … 7](#_Toc27141541)

[4.3 … 7](#_Toc27141542)

[4.4 Bibliography 7](#_Toc27141543)

[5 Data handling 8](#_Toc27141544)

[5.1 Objectives 8](#_Toc27141545)

[5.2 … 8](#_Toc27141546)

[5.3 … 8](#_Toc27141547)

[5.4 Bibliography 8](#_Toc27141548)

[6 Data sharing practices 9](#_Toc27141549)

[6.1 Objectives 9](#_Toc27141550)

[6.2 … 9](#_Toc27141551)

[6.3 … 9](#_Toc27141552)

[6.4 Bibliography 9](#_Toc27141553)

AI4H Data Specification (Outline)

# Objectives

The ITU/WHO Focus Group on Artificial Intelligence for Health (AI4H) has proposed a list of deliverables at meeting "G" in New Delhi in November 2019. The present document proposes an outline for the future deliverable "Data Specification". This deliverable combines a set of six separate deliverables as umbrella, which address six important aspects related to data specification when used for artificial intelligence (AI) and machine learning (ML) models/methods for health purposes (cf. Table I).

Table 1 – Proposed deliverable titles with initial editors and document numbers

| Deliverable title | Initial editors | Number |
| --- | --- | --- |
| Data Specification | Marc Lecoultre | G-205 |
| Data requirements | Gupta Saurabh, Manjula Singh | G-205-A01 |
| Data acquisition | Rajaraman Subramanian, Vishnu Ram | G-205-A02 |
| Data annotation specification | Shan Xu, Harpreet Singh | G-205-A03 |
| Training and test data specification | Luis Oala, Pradeep Balachandran | G-205-A04 |
| Data handling | Marc Lecoultre | G-205-A05 |
| Data sharing practices | Ferath Kherif, Banusri Velpandian | G-205-A06 |

Each editor will propose an initial outline (=Table of Contents), define the objectives of the future deliverable, and collect a bibliography of existing literature and material relevant for the development of the respective document. A short call for participation, the expertise profile of potential contributors, a time plan, and a brief characterisation of the target audience serve as preface.

* Call for Participation
* Expertise profile of potential contributors
* Time plan (first draft, release 1)
* Target audience
* Initial outline
* Objectives
* …
* Bibliography

*The remainder of this document suggests a coherent structure for the deliverables A01-A06,   
which the respective editors may use - if they think it is helpful.*

# Data requirements

Call for Participation

If you are interested in contributing to the *Data requirements*, please contact the editor of this document (Gupta Saurabh, [drsaurabhmd@gmail.com](mailto:drsaurabhmd@gmail.com); Manjula Singh, [singhmanjula.hq@icmr.gov.in](mailto:singhmanjula.hq@icmr.gov.in)) and the Secretariat of the Focus Group ([tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int)) using "ITU/WHO Deliverable - Data requirements " as e-mail-subject, briefly introduce yourself, describe your relevant expertise, and explain your interest.

Expertise profile of potential contributors

To be defined (TBD).

Time plan (first draft, release 1)

TBD.

Target audience

TBD.

Initial table of contents

TBD.

## Objectives

***[Suggestion:]***  Define a framework for public healthcare data acquisition and management model based on standard protocol for its easy adoption by any country or international health organizations. This paper assumes basic digitization of electronic health record (EHR) at basic health facilities. There is a gap in developing an integrated and comprehensive framework that addresses the use of EHR in a standardized way for public health, privacy issue by anonymizing patient specific information, fusing multiple records with slight changes in the same information, augmenting a broad spectrum of contextual data, and so on.

## …

…

## …

…

## Bibliography

*TBD. Format: APA bibliography style with DOI links. References like [Name, Year].*

Suggested reading:

# Data acquisition

Call for Participation

If you are interested in contributing to the *Data acquisition*, please contact the editor of this document (Rajaraman Subramanian, [kinnal@hotmail.com](mailto:kinnal@hotmail.com); Vishnu Ram, [vishnu.n@ieee.org](mailto:vishnu.n@ieee.org)) and the Secretariat of the Focus Group ([tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int)) using "ITU/WHO Deliverable - Data acquisition" as e-mail-subject, briefly introduce yourself, describe your relevant expertise, and explain your interest.

Expertise profile of potential contributors

To be defined (TBD).

Time plan (first draft, release 1)

TBD.

Target audience

TBD.

Initial table of contents

TBD.

## Objectives

TBD

## …

…

## …

…

## Bibliography

*TBD. Format: APA bibliography style with DOI links. References like [Name, Year].*

Suggested reading:

# Data annotation specification

Call for Participation

If you are interested in contributing to the *Data annotation specification*, please contact the editor of this document (Shan Xu, [xushan@caict.ac.cn](mailto:xushan@caict.ac.cn); Harpreet Singh, [hsingh@bmi.icmr.org.in](mailto:hsingh@bmi.icmr.org.in)) and the Secretariat of the Focus Group ([tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int)) using "ITU/WHO Deliverable - Data annotation specification" as e-mail-subject, briefly introduce yourself, describe your relevant expertise, and explain your interest.

Expertise profile of potential contributors

To be defined (TBD).

Time plan (first draft, release 1)

TBD.

Target audience

TBD.

Initial table of contents

TBD.

## Objectives

TBD

## …

…

## …

…

## Bibliography

*TBD. Format: APA bibliography style with DOI links. References like [Name, Year].*

Suggested reading:

# Training and test data specification

*Call for Participation*

If you are interested in contributing to the *Training and test data specification* document, please contact the editors of this document (Luis Oala, [luis.oala@hhi.fraunhofer.de](mailto:luis.oala@hhi.fraunhofer.de); Pradeep Balachandran, [abn.tvm@gmail.com](mailto:abn.tvm@gmail.com)) and the Secretariat of the Focus Group ([tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int)) using "ITU/WHO Deliverable - Training and test data specification" as e-mail-subject, briefly introduce yourself, describe your relevant expertise, and explain your interest.

Expertise profile of potential contributors

To be defined (TBD).

Time plan (first draft, release 1)

TBD.

Target audience

TBD.

Initial table of contents

TBD.

## Objectives

Review of training and test data specification practices, especially those relating to the technical requirements of data for training and testing an AI.

## …

*…*

## …

*…*

## Bibliography

*TBD. Format: APA bibliography style with DOI links. References like [Name, Year].*

Suggested reading:

# Data handling

Call for Participation

If you are interested in contributing to the *Data handling*, please contact the editor of this document (Marc Lecoultre; [ml@mllab.ai](mailto:ml@mllab.ai)) and the Secretariat of the Focus Group ([tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int)) using "ITU/WHO Deliverable - Data requirements " as e-mail-subject, briefly introduce yourself, describe your relevant expertise, and explain your interest.

Expertise profile of potential contributors

To be defined (TBD).

Time plan (first draft, release 1)

TBD.

Target audience

TBD.

Initial table of contents

TBD.

## Objectives

***[Suggestion:]*** Outline how data will be handled, once they are accepted. Health data are one of the most valuable and sensitive types of data. Handling this kind of data is often associated with a strict and factual framework defined by data protection laws. It is important to set a strict data policy which will ensure confidence in FG-AI4H not only among contributors, but across all stakeholders. There are two major issues that the data handling policy should address: (a) compliance with regulations dealing with the use of personal health data; and (b) non-disclosure of the undisclosed test data held by FG-AI4H for the purpose of model evaluation.

## …

…

## …

…

## Bibliography

*TBD. Format: APA bibliography style with DOI links. References like [Name, Year].*

Suggested reading:

# Data sharing practices

Call for Participation

If you are interested in contributing to the *Data sharing practices*, please contact the editor of this document (Ferath Kherif, [Ferath.Kherif@chuv.ch](mailto:Ferath.Kherif@chuv.ch); Banusri Velpandian, [banushrir@gmail.com](mailto:banushrir@gmail.com)) and the Secretariat of the Focus Group ([tsbfgai4h@itu.int](mailto:tsbfgai4h@itu.int)) using "ITU/WHO Deliverable - Data sharing practices" as e-mail-subject, briefly introduce yourself, describe your relevant expertise, and explain your interest.

Expertise profile of potential contributors

To be defined (TBD).

Time plan (first draft, release 1)

TBD.

Target audience

TBD.

Initial table of contents

TBD.

## Objectives

TBD

## …

…

## …

…

## Bibliography

*TBD. Format: APA bibliography style with DOI links. References like [Name, Year].*

Suggested reading:

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