INTERNATIONAL TELECOMMUNICATION UNION

TELECOMMUNICATION STANDARDIZATION SECTOR

STUDY PERIOD 2017-2020

FG-AI4H-F-103

ITU-T Focus Group on AI for Health

Original: English

WG(s):	Plenary	Zanzibar, 3-5 September 2019

DOCUMENT

Source:	FG-AI4H
Title:	Updated FG-AI4H data acceptance and handling policy
Purpose:	Admin
Contact:	TSB tsbfgai4h@itu.int

Abstract: The first part of this document lists *acceptance criteria for data* submitted to the FG-AI4H and states the governing principles and rules. These principles are crucial because the core of the benchmarking framework for AI for health methods will be an *undisclosed test data* set – per use case of each topic area to be defined – that will not be made accessible to the AI developers. The second part of this document outlines *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.

1 Rationale and scope

Artificial Intelligence (AI) can help achieving the important objective of ensuring health for everyone in many ways, worldwide, often at reduced costs and enhanced speed. In the case of modern AI, it is important to notice that practitioners, patients and medical device regulators are confronted with a new kind of machine. While mechanical devices, electronics and software tools from the past have been typically designed from fully understood first principles, it is difficult to anticipate the behaviour of modern AI algorithms, because of the enormous complexity of the algorithms, and because the performance depends not only on the learning algorithm, but also on the underlying training data. These properties let the users raise doubts about whether they can trust AI models, when they face critical decisions in the health domain. Crucially, these reasonable doubts cannot be resolved at present, because there are no established ways to assess the quality of AI models for health.

The Focus Group on "Artificial Intelligence for Health" (FG-AI4H) will meet this need by demonstrating how the performance of AI solutions for health can be evaluated in a systematic fashion. For this purpose, a benchmarking framework will be developed in a best practice type of approach for representative use cases. Having successfully demonstrated the benefits of benchmarking for selected representative use cases, will allow for expanding the approach to a wider range of use cases. Exemplary use cases may include AI-based diagnostics, treatment decision making, triage, patient self-management, risk assessment, image segmentation or annotation, early detection, among others. Obviously not all possible use cases can be addressed considering the limited timespan and resources of the Focus Group.

The core of the benchmarking framework consists of *undisclosed test data* sets - per use case of each topic area to be defined – that will not be made accessible to the AI developers. In addition, (relatively small or large sets of) public data may be made available by FG-AI4H. We would like to note that data publication is not essential for the core idea of the benchmarking framework, but merely an optional extra, and that related problems have already been addressed by others before. Data sets are not limited to any modality such as images, time series, laboratory tests, "omics", text, or electronic health records, but a wide variety is welcome. Details of the envisioned benchmarking procedure are presented in the White Paper of FG-AI4H.

The first part of this document specifies the criteria for data acceptance (cf. section 3). Decisions whether to accept or reject submitted data will be taken according to these criteria. The second part of the document outlines how data will be handled, once they are accepted, and states the governing principles and rules (cf. section 4).

For sensible benchmarking, the topic drivers will address the following three dilemmas: (1) Benchmarking is not valid if AI-techniques developed by data donors are tested on their own donated data, because they know the data and associated output variables/labels. (2) Excluding data donors from benchmarking will considerably reduce the willingness to donate data, which are essential for a reasonable evaluation. (3) Having a data pool from several sources and testing each AI-technique only on data from other sources (i.e. testing AI-technique developed by x only on data donated by y and z) may tempt data donors that also develop AI-technology to contribute as "difficult" data (low quality data, wrong annotations,...) as possible to the data pool, in a competitive setting.

2 Terminology

2.1 Terms defined in this document

In this document, we refer to different types of datasets. For clarity, we suggest the following definitions:

Received data: Any dataset submitted by a trusted source (tbd) and received by FG-AI4H;

Public data: Subset of the *received data* that is made public by FG-AI4H to help AI developers to understand the structure of the undisclosed test data, or to train AI technology if enough data are provided;

Undisclosed test data: Corresponds to the remaining *received data* after removing *public data*. This set is kept strictly private to evaluate submitted AI technology.

2.2 Specific terms

When we use the terms "shall", "should" and "may", they have a specific meaning which is explained in the next table:

Term	Meaning
"Shall"	states a **mandatory** requirement of this policy
"Should"	states a **recommended** requirement of this policy
"May"	states an **optional** requirement

3 Criteria for data acceptance

3.1 Mandatory

The data set and the targeted use case are described clearly and concisely. The use case is relevant and of interest for FG-AI4H (selected after prioritization among the various possible use cases). The

data acquisition procedure is described in detail, such that experts from independent trusted institutions can acquire more *undisclosed test data* according to this protocol.

The data type (e.g. images, time series, laboratory tests, "omics", text, electronic health records, etc. or combinations thereof), size (e.g. file size, number of samples), structure (e.g. database type, file format and content etc.), and properties (depending on the data type) are indicated. Any data (pre-)processing methods are explained: It is explained how missing, uncertain, or incomplete data have been treated if they occur. (E.g.: are there any gaps or redundancies, if the submitter provides time series patient or clinical data - in the sense of data with a unique identifier collected over certain time intervals, not continuous EKGs data type? Are imputations or projections of the data available?)

Have the raw data been preserved or have the submitters applied any cleaning mechanism or transformation on the collected data? The data provenance/source is named: Who has collected and/or aggregated the data and where? Who has created the labels/ground truths? Who has assessed the data, e.g. with respect to quality? What were the objectives of the data acquisition? What is the current ownership of the data? Data and annotations/labels/output variables have been validated by an independent domain expert/specialist in terms of quality and suitability, especially in the case of automated data annotation procedures.

The data follow the applicable laws and regulations for data acquisition, processing and sharing, such as privacy laws, copyright laws etc. Contact details and relevant information about the submitter are given. Any potential conflicts of interest are clearly indicated.

The *undisclosed test data* are crucial for the benchmarking procedure. Therefore, the safe storage has to be assured (cf. section 4 on data handling). The measures that guarantee secrecy are described and it is specified who has had access to the *undisclosed test data* in the past, at present, in the future (e.g. published or plan to share with other researchers). How and where are the *undisclosed test data* currently hosted/stored? Consent is given to keep the *undisclosed test data* undisclosed. Clearance is demonstrated for the use in benchmarking (under compliance with the relevant laws, e.g. copyright, privacy). The *undisclosed test data* are suited for benchmarking (to be defined by respective working groups and topic drivers).

3.2 Conditional

If the data originate from humans or are related to humans, the principles of the Declaration of Helsinki have to be adhered to [World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013; 310(20):2191-2194. https://doi.org/10.1001/jama.2013.281053]. Approval was obtained from the local ethics committee where the data were generated (if applicable). Informed written consent to data acquisition, processing and sharing was obtained from the respective person. The anonymization/ pseudonymization and privacy procedure is detailed, and follows the best practices from hospitals or other institutions.

If a subset of the submitted data will be published (public data). Clearance is demonstrated for data publication under compliance with the relevant laws.

Input and output variables are characterized (with codes, classifications, triage tags, pixel or voxel labels, annotations, where they exist in the dataset). Whenever applicable, such characterization shall be conformed to existing health data standards, with the understanding that local or regional extension, restriction, profiling or adaption may be applied.

3.3 Recommended

Submitters transparently describe potential biases. (Bias can - arguably - not be avoided in typical cases of data acquisition and can be expected even in an expert setting. In hospitals, diagnoses and treatment decision are made by experts but might be biased towards reimbursement from health insurances.)

The data qualifiers are described (degree of measurement precision, definition of the quality standards). The data are of sufficient size to create a statistically valid output report. (Otherwise, further data donors need to be found and added.) Further criteria of the data (heterogeneity, real-world relevance etc.) might be considered depending on the use case (to be defined by respective working groups and topic drivers).

It is described how the data can be loaded. (Special software required? Data loader/importer functions available for common programming languages? In case of API access to the data, are there any limitations of the APIs in terms of response time or size of data packages the API endpoints are expected to return? How is the release of a new API version handled? If the APIs are used directly in the tests, will a new release maintain backwards compatibility? This information ensures that the tests will not break.)

Can the submitter help to record more data, in principle? Are the data comparable to other similar data sets? What are the submitter's data handling procedures and data governance processes? How does the submitter handle data versioning? What is the frequency of data updates, if applicable to the particular use case?

4 Data handling

Understanding the importance of data to our initiative and how that information is handled reflects our commitment as a secure organization. The purpose of a data handling policy is to ensure that all sensitive data are confidentially controlled whether being transmitted within the organization or to a trusted third party.

When handling data, all users should be in accordance with and be responsible for adherence to strict and rules to be defined in a reference document. Periodic auditing of adherence to this policy shall be the responsibility of one Focus Group.

Data should be handled in the context of a multi-tiered security system that safeguards patient data according to government statute and regulations. Data should be hosted in secured data centres.

The system shall comply with all applicable regulations over the targeted countries (EU regulations, GDPR, US HIPAA, individual countries healthcare privacy regulations, etc.). Regulations include information security, privacy and quality laws, guidelines and standards. We should design a regulatory compliance framework to ensure conformance with these regulations.

4.1 Legal context

Where national data protection laws may differ significantly, it is important to cover the most restrictive matters to allow the greatest number of entities to share their datasets. This includes data security, anonymization, access control and many other matters discussed in this document.

4.2 Data security

The infrastructure for data storage and processing should be based on state-of-the-art security policies, practices and located in a secure location. Information should be securely received, stored and transferred. Encrypted transmission of datasets and encryption at rest (data stored encrypted) are among many other requirements. Only well-established and approved by FG-AI4H transfer methods should be used (TBD).

Where possible, data transfers should be carried out, using existing, protected and trusted networks (internal to FG-AI4H or over Virtual Private Network with dedicated IPSec & SSL encrypted channels). However, there may be occasions where data will need to be transferred via other networks such as Internet or any other open networks. On these occasions, the data files should be protected by encryption to prevent usage by unauthorised parties.

In case of a physical data transfer, e.g. USB or hard disks, all data should be securely stored in an encrypted format using a method approved by FG-AI4H. Transfers of data in hard copy format should be protected, using methods such as approved secure couriers.

4.3 Data integrity

Data integrity should be enforced when the data travels from one component to another using checksum mechanisms that guarantee that the data have not been corrupted or modified. Any data files transferred or generated should be digitally signed and the data integrity of the payload should be validated at the edge of the network prior to storing the data in the database. This would ensure validation of data integrity of all raw and interpreted patient data.

Any corrupt data (inaccurate or incomplete) should either be rejected by the system or removed from it.

The security & privacy architecture should be designed to ensure a high level of data integrity and privacy for Protected Health Information in compliance with GDPR, US HIPAA, or any other participating country healthcare privacy, security, and quality regulations. This may be dependent on where the data was transferred from, where the data will be processed and by which entity.

4.4 Access control

Authorised stakeholders need to access the data for their own defined purpose and infrastructure administrators for maintenance. The receiving parties such as the Working Groups should evaluate and work on the datasets. The organizations that are willing to submit their algorithms need to access the *public data* to develop their models. To guarantee absolute fairness among submitting organizations and ensure the credibility of the Focus Group, the *undisclosed test data* should remain undisclosed.

Clear access control should be defined and a database with detailed access rights policies should be implemented.

The system should authenticate users before any access to the system and its resources. The system should support standard authentication technique that can verify the identity claimed by the user (Claims based, Federated authentication…).

Everyone willing to submit an algorithm should have access to the *public data*. The only restriction might be for the party submitting the *undisclosed test data*.

4.5 Auditing / Logging

All transactions should be authenticated, authorized, monitored, and logged and audited regularly to detect unauthorized events. The system should detect events that can affect the confidentiality of personal health data or content of the *undisclosed test data*. The system should also record a trail of all processing of personal health information or *undisclosed test data*, such as viewing, creation, modification, validation, printing, copying, import, export, transmission, reception.

Unauthorized access attempts should be denied, and all requests should be logged and retained for audit purposes. Audit logs should be stored in encrypted form and decrypted only by recorded authorized requests and analysed as potential breaches.

4.6 Data lifecycle

Data lifecycle reflects all the steps and the related data processing and management capabilities followed by data from its creation to its use and disposal. From a data point of view, the listed capabilities might affect the state and structure of data, the location of the data, its combination with other data, its transformation, its use and its disposal.

Processing and managing data require effective data governance. Data governance refers to the overall management and caretaking of data, from creation to deletion, covering usability integrity and security. The data governance process should be defined to determine what data is retained or deleted. Data should be kept, so in the case of the creation of a new benchmark, models could be retested.

Once the data is received, it should be stored in a temporary location until data quality validation (verification or detection of any data abnormalities) is completed before transfer to the production environment.

When required, data should be securely erased in accordance with a data destruction policy.

4.7 Data processing

Data processing is the ability to handle data as input and apply different treatment that might modify the data, or combine it without modifying it with other data in order to produce an output that is useful for a given application or service in the data lifecycle.

During evaluation phase, *undisclosed test data* needs to be decrypted. We should ensure non-disclosure of the data during this critical phase.

4.8 Data ownership

The use and ownership of *Received data* should be clearly defined in a licence agreement between the party providing the data (the owner of the data) and the FG-AI4H.

4.9 Backup and archiving

Backed-up or archived data should have at least the same level of protection as those in use, it should be encrypted. Backup should be in separate secure location.

4.10 Interoperability

In case we foresee any need for interoperability with other health institutions or participation in any Open Data initiatives, we might have to decide on a Data Warehouse/Registry structure and how to build standardization around the data.

We might provide APIs/Web services to open different data exchange channels for collaboration with other partners.

4.11 Compliance with international standards

Yearly audits should be conducted by internationally accredited auditors to confirm ITU/WHO observe obligatory security, data protection, continuity and compliance guidelines and procedures. This could comply with international standards such as ISO 27001.

The security architecture for the Data Repositories should comply with security policies and privacy policies. The security solutions should be in alignment with ISO 7498-2 Security Model best practice recommendations on information security management.

4.12 Risk assessment

There should be periodic assessment of the potential risks and vulnerabilities to the confidentiality, integrity, and availability of electronic protected information held in the repository.

We should conduct a proactive periodic risk analysis of the audit logs and should take corrective action when unacceptable risks are identified. Proactive security measures sufficient to reduce risks and vulnerabilities to the level required by the data’s high sensitivity shall be maintained throughout the programme lifecycle.

_____