



FG AI4H会合報告

慶応義塾大学大学院 政策メディア研究科 特任教授

かわもり まさひと
川森 雅仁



1. はじめに

2018年7月に始まったITUとWHOが共同で開催しているFocus Group on AI for Health (FG-AI4H) は、2020年のコロナ禍が発生して以降、オンライン開催になっており、2021年の最初の会合は1月27日から29日 (K会合) と5月19日から21日 (L会合) に開催された。2021年に入って、FGとしても最終的な成果に向けてかなりまとまった結果が見えてきた。

1.1 FGの参加状況

2021年最初の2会合は、SG16の会合を間に挟んだことから、FGの一区切りにあたり、まとめのレポートが出された。FGの参加状況に関する報告もあり、現在までの各国の参加状況は表のとおりである。

現在までの延べ人数は95人であり、国としては31か国が参加している。ドイツ、スイス、中国、インド、英国、米国の6か国で70%を占めている。また、アフリカから5か国 (エジプト、カメルーン、ガーナ、タンザニア、ウガンダ)、東南アジアから3か国 (インドネシア、シンガポール、タイ) が参加した。

2. 特筆すべき結果

約3年を経て、今回の会合はFGにとっての一区切りであり、出力文書の体系化、今後の方向性が改めて明確化された。

またコロナ禍に対応するためのガイドラインの作成も特筆される。

2.1 COVID-19ガイドラインの作成

FG-AI4Hは、新型コロナウイルスの全世界的な蔓延による緊急事態のためのAI及びその他のデジタル技術を使用したベストプラクティスを集めたレポートを作成した。これには、予防と準備、発生の早期発見、監視と対応、復旧、リハビリ、緩和などが含まれている。この点に関しては、国連の「仙台防災枠組 2015-2030」を参照している。この技術文書はコロナ禍だけでなく今後起こるかもしれない、健康上の緊急事態に適用可能であることが期待されている。

2.2 Open Code Initiative (OCI) の形成

FG AI4Hで今まで議論してきたベンチマークプラットフォームとしてまとめられたフレームワークは既に一部実装されており、AI4Hのオープンコードソフトウェアツール (FG-AI4H OCI) として利用可能にした。AI4Hは、オープンコードイニシアチブの一連のツールの開発とテストにさらに役立つことが期待されている。

3. 成果文書案

FG AI4Hの成果文書 (Deliverables) 体系を図1に示す。ここから分かるように、重要な核をなす文書群はDEL01の

■表. FGの各国の参加状況

番号	国名	人数	番号	国名	人数	番号	国名	人数	番号	国名	人数
1	Germany	15	11	Austria	1	21	Indonesia	1	31	Uganda	1
2	Switzerland	14	12	Brazil	1	22	Italy	1			
3	China	12	13	Cameroon	1	23	Japan	1			
4	India	10	14	Canada	1	24	S. Korea	1			
5	UK	10	15	Colombia	1	25	Netherlands	1			
6	USA	6	16	Czech Republic	1	26	New Zealand	1			
7	Bangladesh	2	17	Egypt	1	27	Tanzania	1			
8	Portugal	2	18	Ghana	1	28	Thailand	1			
9	Spain	2	19	Estonia	1	29	Nicaragua	1			
10	Australia	1	20	France	1	30	Singapore	1			

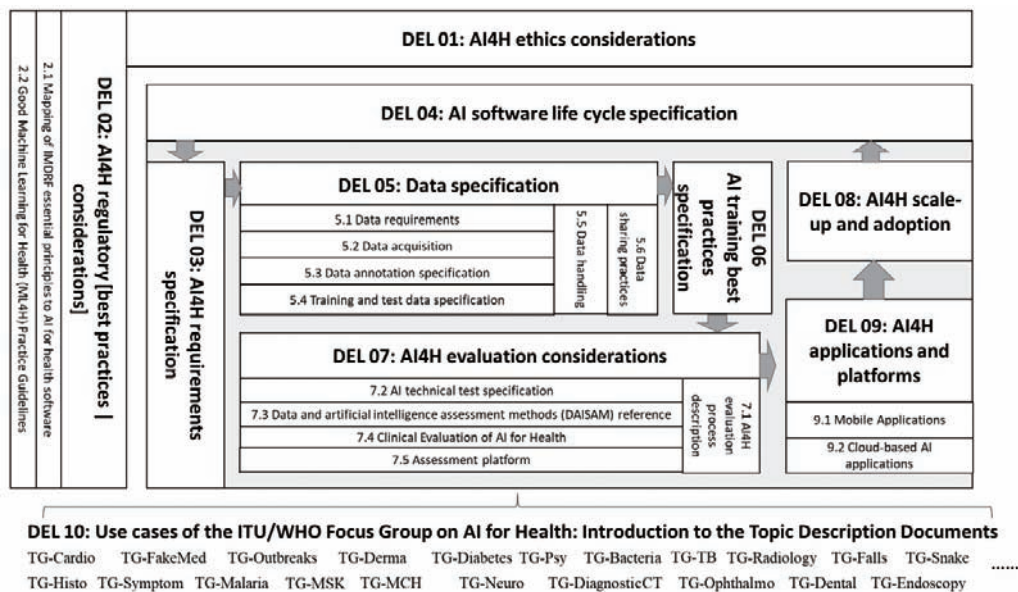


図1. FG-AI4H出力文書体系

AIの倫理的考慮、DEL04のライフサイクル定義、DEL05のデータ仕様、DEL07の評価法であるといえる。DEL02の制度のベストプラクティスは、各国によって様々な違いがあり、参考的扱いになっている。

またDEL05は、今回の会合でOCIとしてパッケージ化されることになっている。

またDEL10のユースケースは実際にAIを使った例とデータを集めているという点で、FG-AI4Hの最大の成果といえるかもしれない。

DEL-01 AI4Hの倫理に関する考慮事項 (AI4H ethics considerations)

AI技術の倫理的開発と使用に関する多くの課題は未解決のままであるが、特にWHOは、このような新技術をもたらす倫理的課題について倫理的、規制、専門的、経済的影響を理解して積極的に取り組む責務を負っている。ゆえに、この文書はWHOが中心で作成されている。今回もいくつかの問題提起がなされたが、文書化は進んでいない。現在の文書案はTD K-028に収められている。

DEL-02 AI4H規制のベストプラクティス (AI4H regulatory best practices)

この文書も上記の倫理に関する文書同様、WHOが中心になって作成しているAI4Hの規制に関するガイドラインである。図にAI機器のライフサイクルに関する規制の在り方

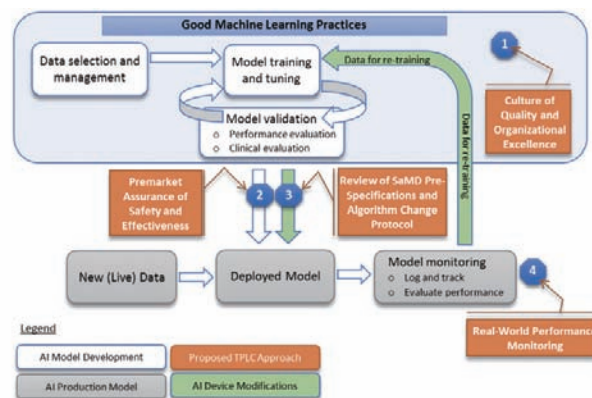


図2. AI機器のライフサイクル

の例を挙げる。

現実的には医療や健康に関する規定は各国で定められるものであるため、このベストプラクティスも参考文書程度にしかなり得ない。あくまで、健康分野にAI4Hの結果を利用しようとした場合のAI開発側と規制当局間の共通理解を確立するための対話を促すに資することが期待されている。

DEL02-1 IMDRF基本原則のAI4Hソフトウェアへのマッピング (Mapping of IMDRF essential principles to AI for health software)

国際医療機器規制当局フォーラム (IMDRF) は世界各国の医療機器規制当局による任意活動で医療機器規制の国際整合化について将来の方向性を議論するフォーラムと

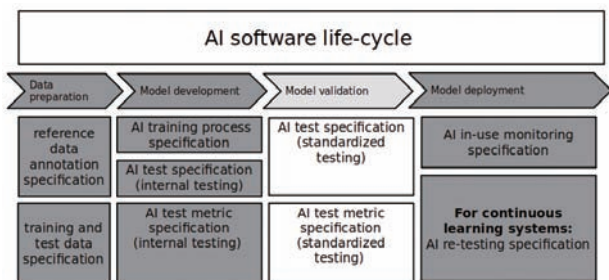


して2011年2月に設立された。国際的な医療機器規制の整合化と収束を促進することを目標としている。本出力文書は、IMDRFが発表している「医療機器としてのソフトウェア」(Software as a Medical Device: SaMD) の規制枠組みとして“Essential Principles of Safety and Performance of Medical Devices and IVD Medical Devices”, (2018) とAI4HのAIソフトウェアの間の対応関係をまとめた資料である。この資料は、2019年の11月に第一版 (FG-AI4H-G-038) がまとめられたとされ、その後進展がないので、これが最終版になると思われる。

DEL02-2 AIベースの医療機器のガイドライン—規制要件

この文書は、規制当局及び医療機器メーカーに対して医療機器向けAI (AI-MD) が順守すべき包括的要件を分析可能にするベストプラクティスを示し、その工程についての共通理解を持つための一連のガイドラインを定義している。出力文書としては1月会合の出力文書FG-AI4H-K-039に規定された。

この文書のガイドラインは、包括的なものではなく、またあらゆる国、領土、地域における、規制、法律、標準等の代わりとなることを目標としていない。



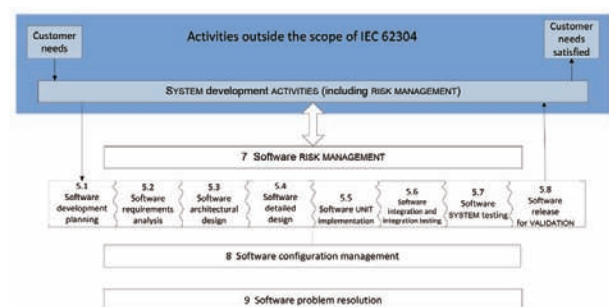
■図3. AIソフトのライフサイクル

DEL03 AI4H要件仕様 (AI4H requirement specifications)

この文書は、一般的なAI4Hシステムの情報、機能、動作及び運用等を記述するシステム要件仕様 (SyRS) を定義している。これは、AI一般のシステム要件を記述しているのではなく、AI4Hのプラットフォーム上で使用されるための要件を記述している。例えば、DEL10のユースケースのために用いられるシステムはこの要件に従っていることが想定されていると思われる。しかし、実際のところ適合性が規定されていないため要求条件に合っているかどうかを判断するのは比較的困難だと思われる。現状版はFG-AI4H-K-040として出力されている。

DEL04 AI4Hソフトウェア・ライフサイクル仕様 (AI software life cycle specification)

この文書はAI4Hソフトウェアのライフサイクルに関連する標準とベストプラクティスを特定し、他のソフトウェアのライフサイクルと同様に、AIソフトウェアのライフサイクルを指定する、ということになっている。しかし、実際には、特に医療機器の観点から、ライフサイクルの基本的定義として、IEC 62304 “Medical device software: software lifecycle process” を挙げている。



■図4. IEC 62304の構成

これを補完するものとしてIEC 82304-1 “Health software-Part 1: General requirements for product safety” を挙げており、基本的にこれらの外部標準を参照することでAI4Hのライフサイクルとする。また特に医療機器のリスクマネジメントとしてISO 14971:2019, “Medical devices-Application of risk management to medical devices” を参照している。

出力文書は前回会合のFG-AI4H-J-033であり、不完全ではあるがこれ以上の進展があるかどうかは不明である。

DEL05 データ仕様 (Data specification)

DEL05文書は、AI4Hが扱うデータに関連するデータに関する要件、アノテーション、収集、訓練とテスト、処理、共有の6分野を記述した文書からなる。

DEL05.1 データ要件 (Data requirements)

この文書には、FG-AI4Hに送付されたデータの受け入れ基準がまとめられており、その基準に準拠するための原則と規則が記載されている。AI4Hのベンチマーク枠組みの核となるのは、AI開発者がアクセスできない、各特定医療領域 (後に述べるTGの領域) のユースケースごとのテスト用非公開データセットであるので、これらの原則は非常に重要である。



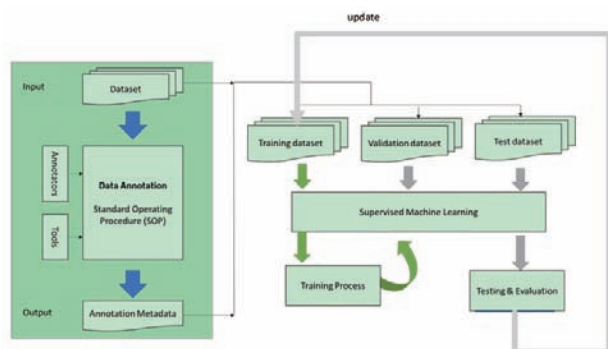
また個人情報について、1964年に世界医師会総会で採択された「ヒトを対象とする生物医学的研究に携わる医師のための勧告」に言及して、プライバシー保護の徹底が要求されている。

DEL05.2 データ収集 (Data acquisition)

この文書は、どのような国や国際的医療機関でも簡単に採用できるように、標準プロトコルに基づく公的医療データの取得及び管理モデルのフレームワークを記述している。中核となる概念は、データレイク (Data Lake) である。データレイクとは、「湖」という言葉の比喻のように、様々なソースからのビッグデータを加工・変換などせずに元の形式のまままで保持する広大な領域のことである。現状の版は、かなり前のFG-AI4H-G-205-A02であり、その後変更が加えられていない。

DEL05.3 データ・アノテーション仕様 (Data annotation specification)

本文書は、データ・アノテーション仕様の一般的なガイドラインである。FG-今回の会合の出力文書AI4H-K-048に収められている。



■ 図5. データアノテーションと他機能との関係

DEL05.4 トレーニングとテストデータ仕様 (Training and test data specification)

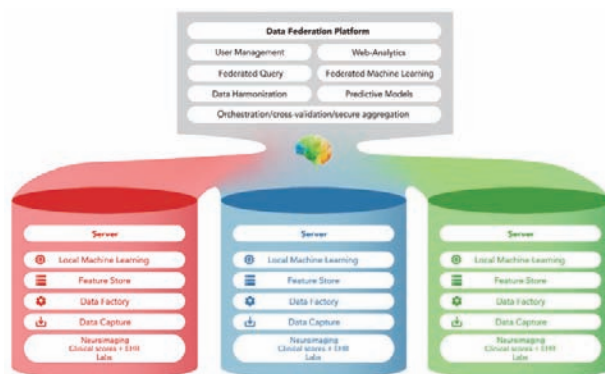
この文書は、機械MLモデルのトレーニングとテストで使用するデータセットの技術要件仕様を体系的に記述する。出力文書としては、前々回会合でまとめられたFG-AI4H-I-034をもとにして、本会合からはOpen Code Initiative (OCI) Data Acquisition and Storage Package (DASP) へと移行することになった。(FGAI4H-K-050)

DEL05.5 データ処理 (Data handling)

この文書では、データが受け入れられた後のデータ処理方法について概説することになっているが、まだ文書としてまとまっていない。現在、最終版とされている、FG-AI4H-DEL05に含まれているのは、別内容である。

DEL05.6 データ共有の実践 (Data sharing practices)

この文書は、健康関連データのデータ共有に関する現存するベストプラクティスの概要を記述し、安全なデータ共有を可能にするために必要な手順と要件を記述することになっている。既存の分散環境とフェデレーション環境に基づく方法とより新規性のあるアプローチが記載されている。

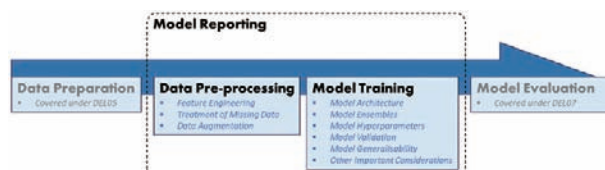


■ 図6. データフェデレーションの例:

この文書では、DAQCORD (Data Acquisition, Quality and Curation Observational Research Design) の利用が提案されている。DAQCORDは、ケンブリッジ大学などで開発された生物医学分野での大規模な観察研究のためデータ品質指標を集めた包括的なガイドラインで、神経科学分野を中心に開発されたものだが、他の健康研究分野や、小規模な前臨床研究にも関連性があり、一般化できるとされる。

DEL06 AIトレーニングのベストプラクティス仕様 (AI training best practices specification)

この文書は主にトレーニングされたモデルのレポートに関するベストプラクティスを記述している。出力文書は、G-AI4H-K-037に含まれている。

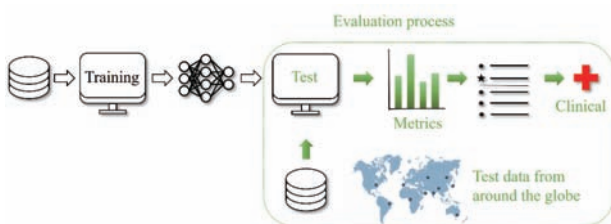


■ 図7. AIトレーニングとレポートの関係



DEL07 AI4H評価事項 (AI for health evaluation considerations)

この文書は、AI4H評価事項に関わる、DEL07シリーズの6つの文書の概要とその背景情報を提供する。特に重要な評価の考え方は、信頼できる第三者による、様々なデータソースから取得した高品質テストデータに対して標準化されたテスト手順と評価法を備えたベンチマークを構築する、ということである。



■図8. 標準的ベンチマークのイメージ

このベンチマークでは、AI4Hでの評価プロセス記述、技術試験仕様、測定基準、臨床評価などが記述されることになっている。

今回新たに評価プラットフォームが出力として加えられた。

DEL07.1 AI4H評価プロセス記述 (AI4H evaluation process description)

この文書は、WHOのエディターによって、最先端のAI評価のための原則と方法の概要を記述しREAL AI概念へのアプローチを記述することになっている。しかし、現在のところ文書案は、目次が挙がっているだけで、この先どのように作成されるか注目される。

DEL07.2 AI技術試験仕様 (AI technical test specification)

この文書では、コンピュータによる (in silico) AIの試験法とその要求条件を規定している。検討されている試験法には、以下のようなものがある。

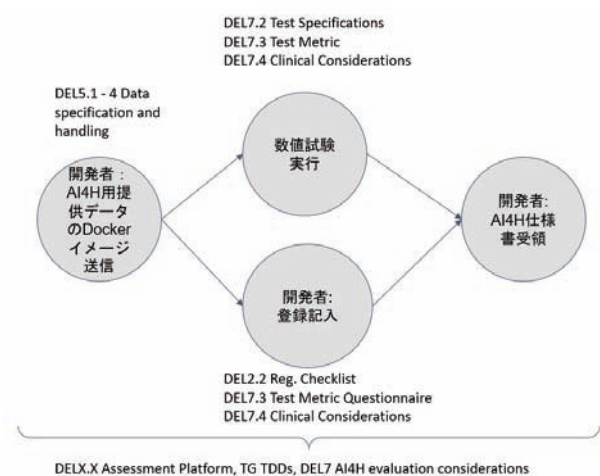
機能試験、非機能試験、ホワイトボックス試験、ブラックボックス試験、メンテナンス試験、静的試験、変更関連試験、破壊試験。現在、文書化はあまり進んでおらず、前々回の会合から文書はアップデートされていない。

DEL07.3 データと人工知能の評価方法 (DAISAM) リファレンス (Data and artificial intelligence assessment methods (DAISAM) reference)

本文書は、アルゴリズム上の偏り、解釈可能性、堅ろう性、

不確実性など、データとモデルの品質の評価方法を定義する参照文書という位置付けである。

DEL07の文書群はDAISAM (Data and artificial intelligence assessment methods) と呼ばれているが、その中核となる文書とされている。ただし、これは汎用的な試験を定義しているわけではなく、AI4Hに提供されるデータ等の品質を保つための基準を示しており、一般的なAIやMLのデータに適用されるものではない。図9に試験要領概略を示す。



■図9. AI4Hの試験要領概略

ここから分かるように、データは、Docker (コンテナ仮想化を用いてアプリケーションを開発・配置・実行するためのオープンソースプラットフォーム) を使って提供される。

DEL7.4 AI4H臨床評価 (Clinical evaluation of AI for health)

FG-AI4Healthの目的の一つは、規制当局や意思決定者に対して、AI4Hの「最新バージョン」の標準が実運用状況やユースケースに適していることが明確になるような中間ステップを、速やかにかつ信頼のおける形で提供することである。本文書は臨床でのAI使用に関して現状の例のレビューを提供し要求条件を記述しようとしている。他の文書との重複が気になる文書である。

DEL07.5 評価プラットフォーム (Assessment platform)

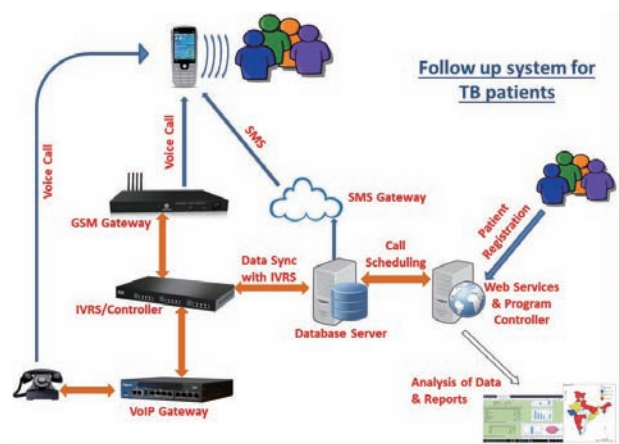
本文書は、標準化された方法で幅広いユースケースのAI4Hモデルの評価のため/ベンチマークプラットフォームの実装するいくつかの方法を記述することになる。

DEL08 AI4Hの拡張と採用 (AI4H scale-up and adoption)

この文書はまだ内容が確定していない。

DEL09 AI4Hアプリケーションとプラットフォーム (AI4H applications and platforms)

この文書は、モバイル及びクラウドベースのAIアプリケーションの開発例やベンチマークの規則案を標ぼうしている。実際には、単に実装例を挙げているに過ぎない。図10に今回議論された使用例を示す。



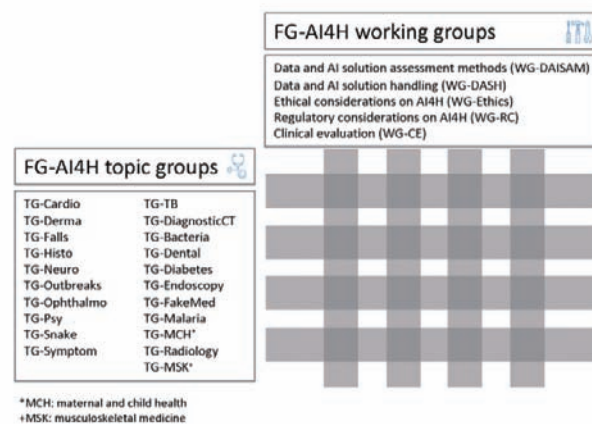
■ 図10. インドでの結核モニタリングの例

これは、インドでの携帯アプリを使った結核患者モニタリングの例であるが、図から分かるように、SMSとVOIPを使って患者が情報にアクセスできる、というものである。インドでは、1億人以上（人口の8.4%）が結核患者ということで重要なサービスといえる。しかし、AIを使っていない普通の電話通信である。FG.AI4Hは、発展途上国からも参加者が多いが、通信インフラ整備の程度によって、AI以前にやるべきことがある国、地域が多い、という例でもある。

4. 特定分野トピック文書案

上記のAI4Hの共通項目に関する出力文書に対して、特定の健康医療領域に関して20のトピックグループが抽出され、お互いに重複を避け、専門性を維持しながら、様々なユースケースを収集し、それに対応するAI/MLベンチマークタスクを記述する。これらは、AI4H use casesとして10番目の出力文書の範ちゅうとなる。

以下にそのトピック分野と他のWorking Groupとの間の関係を示す。



■ 図11. TG文書とWG文書との関係

また今回新トピック分野として以下の3つが加えられた。

- ・ TG-Fertility生殖と出産 Human reproduction and fertility
- ・ TG-POC臨床現場診断 (Point-of care diagnostics)
- ・ TG-Sanitation公衆衛生 (Sanitation for public health)

5. 今後の予定と課題

FGの進捗報告によると、FGの参加者たちは、進捗状況に満足しており、初期の運用を経て作業方法等がより明確になり安定したと思っているため、これを基にFGの根幹的な課題の進捗がさらに見られるだろうという期待を持っている。

一方、文書化は、実際の運用に追いついていない印象で、完成度の低いものも多い。また、AI4H内でのデータのやり取りに関する規定にとどまっているものが多く、結局AI4H内に大量のデータを蓄え、それをオープンソース化したプラットフォームを作ることが目的であるかのような印象も受ける。そうすると、様々な機関で作られているデータベースと同列のものを作ることに過ぎない、とも思える。

また、参加していく組織の研究結果の発表で終わっているものも多い。標準化活動と、どう関連していくかが課題といえる。

2021年4月のSG16会合において、FG-AI4Hはさらに運営を継続することが承認された。

FG-AI4Hの次の会合は、2021年の秋にN会合が開催されることになっているが、日程は未定である。