資料1

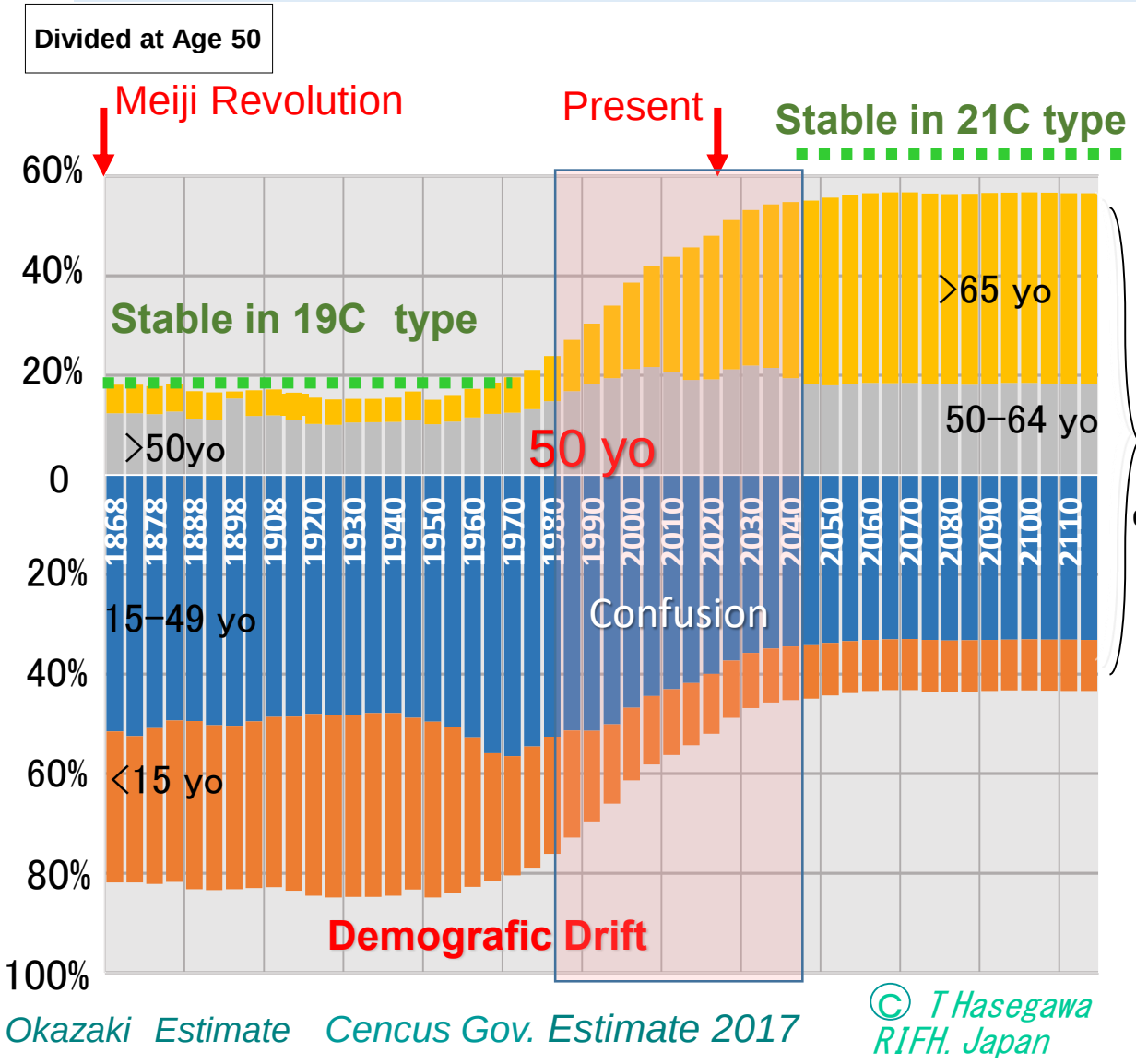
AMEDにおける画像関連データベース および共通プラットフォーム構築研究の現状

２０１９年６月６日（木）

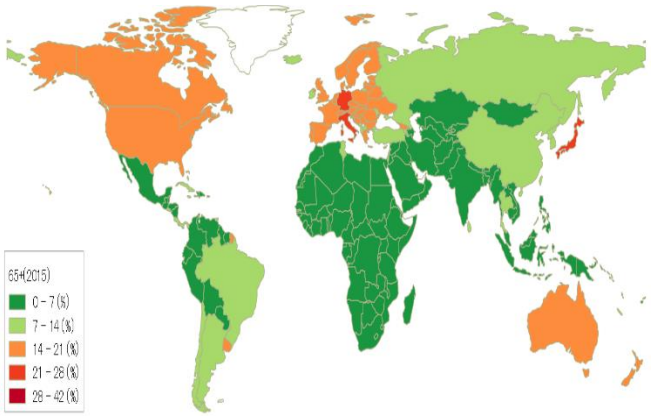
第８回　保健医療分野AI開発加速コンソーシアム

日本医療研究開発機構　理事長　末松　誠

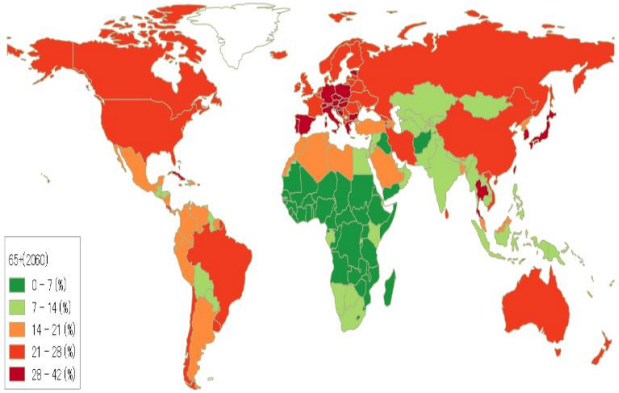
Great Demographic Transition in Japan



Population(%) age over 65 years in 2015

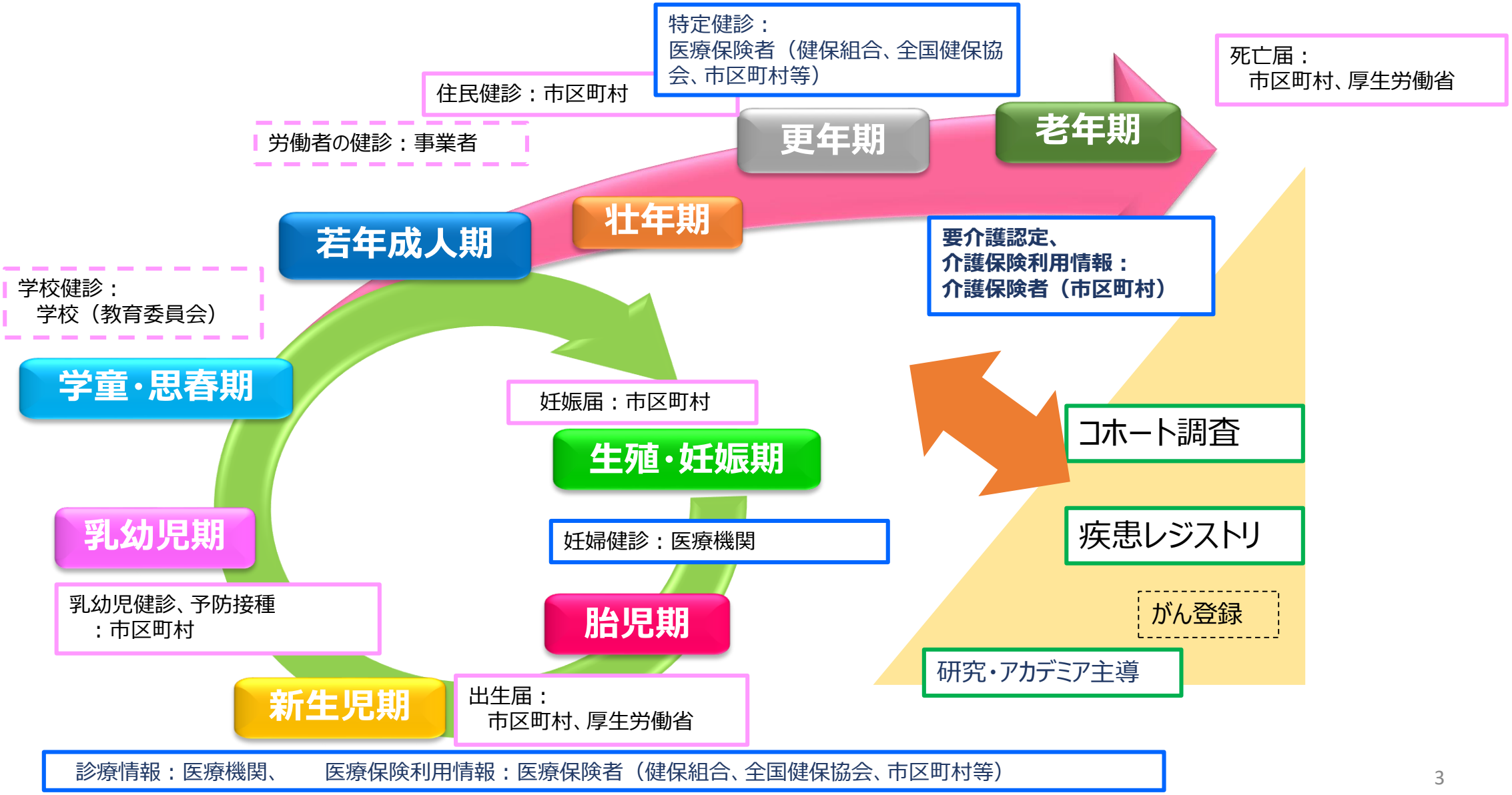


in 2060



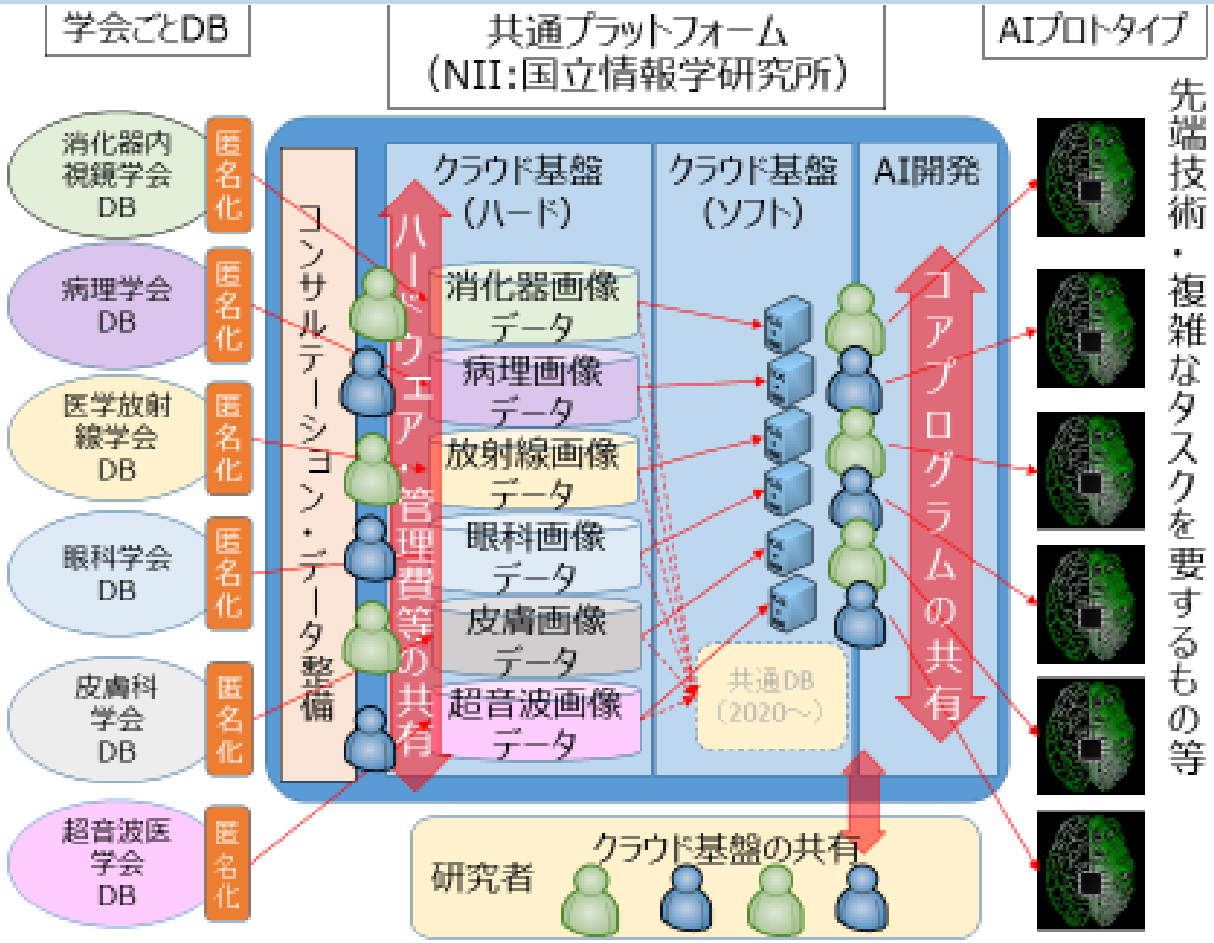
ヒト生涯の健康に関わるデータ（概念図）

Copyright 2015 Japan Agency for Medical Research and Development. All Rights Reserved.



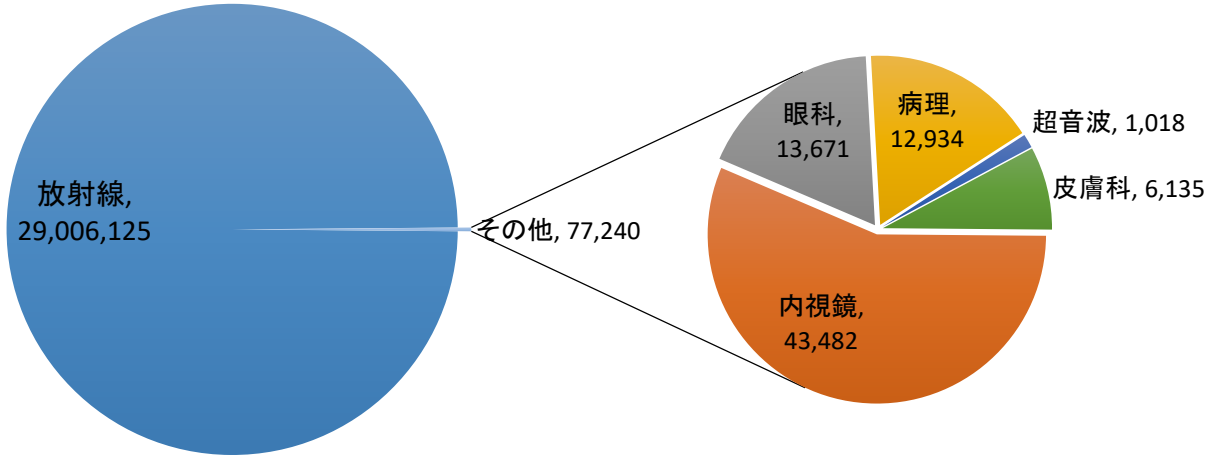
AMEDによるAI関連研究課題の支援

複数の学会連携による競争領域・協調領域の環境整備



クラウド基盤 - 接続・データの利用状況

平成30年度医療画像データ受入実績
(単位：画像の枚数) 2019.4.4現在

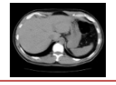


国立情報学研究所（NII）による共通プラットフォームの構築とSINET5を利用した画像情報の集積

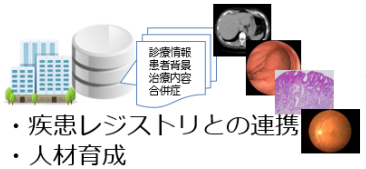
日本病理学会
病理専門医数：
2,232名

日本消化器内視鏡学会
消化器内視鏡専門医数：
16,200名

日本医学放射線学会
放射線診断専門医数：
5,057名



診療画像等データベースプラットフォームの形成



・疾患レジストリとの連携
・人材育成
・広域連携・分散統合型のプラットフォーム

国立情報学研究所（NII）
AI解析等による
診療支援ソフトや
診断補助システムの開発

AI開発基盤の構築

広域連携・分散統合

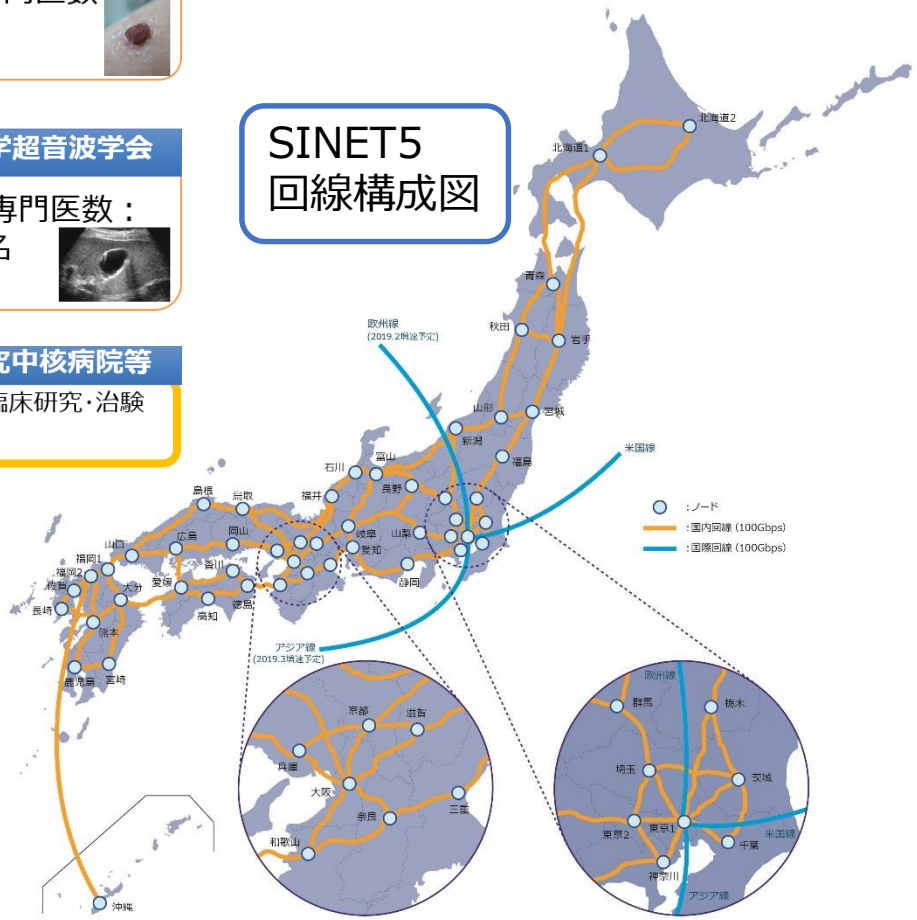
日本眼科学会
眼科専門医数：
10,860名

日本皮膚科学会
皮膚科専門医数：
6,129名

日本医学超音波学会
超音波専門医数：
1,829名

臨床研究中核病院等
効率的な臨床研究･治験への応用

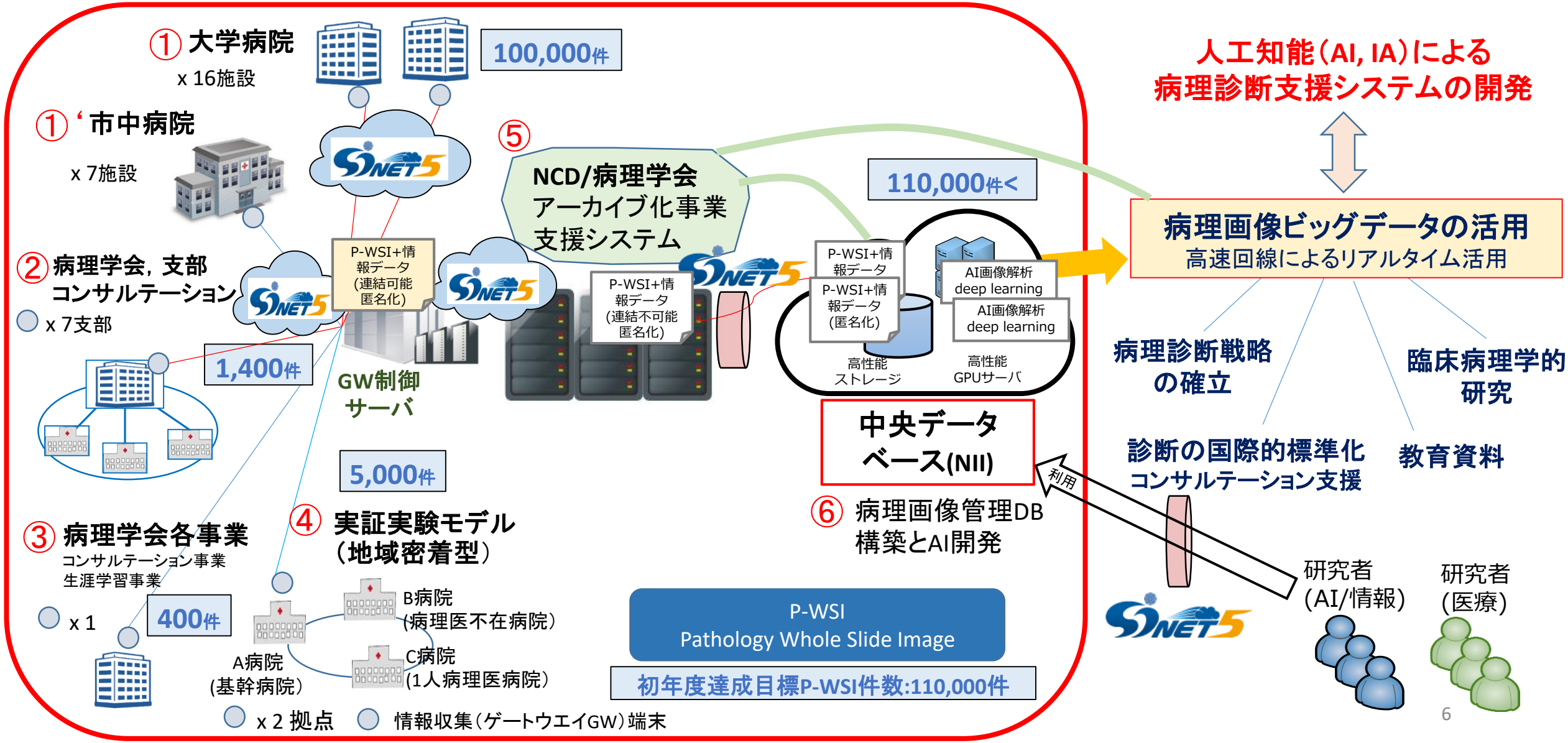
SINET5
回線構成図



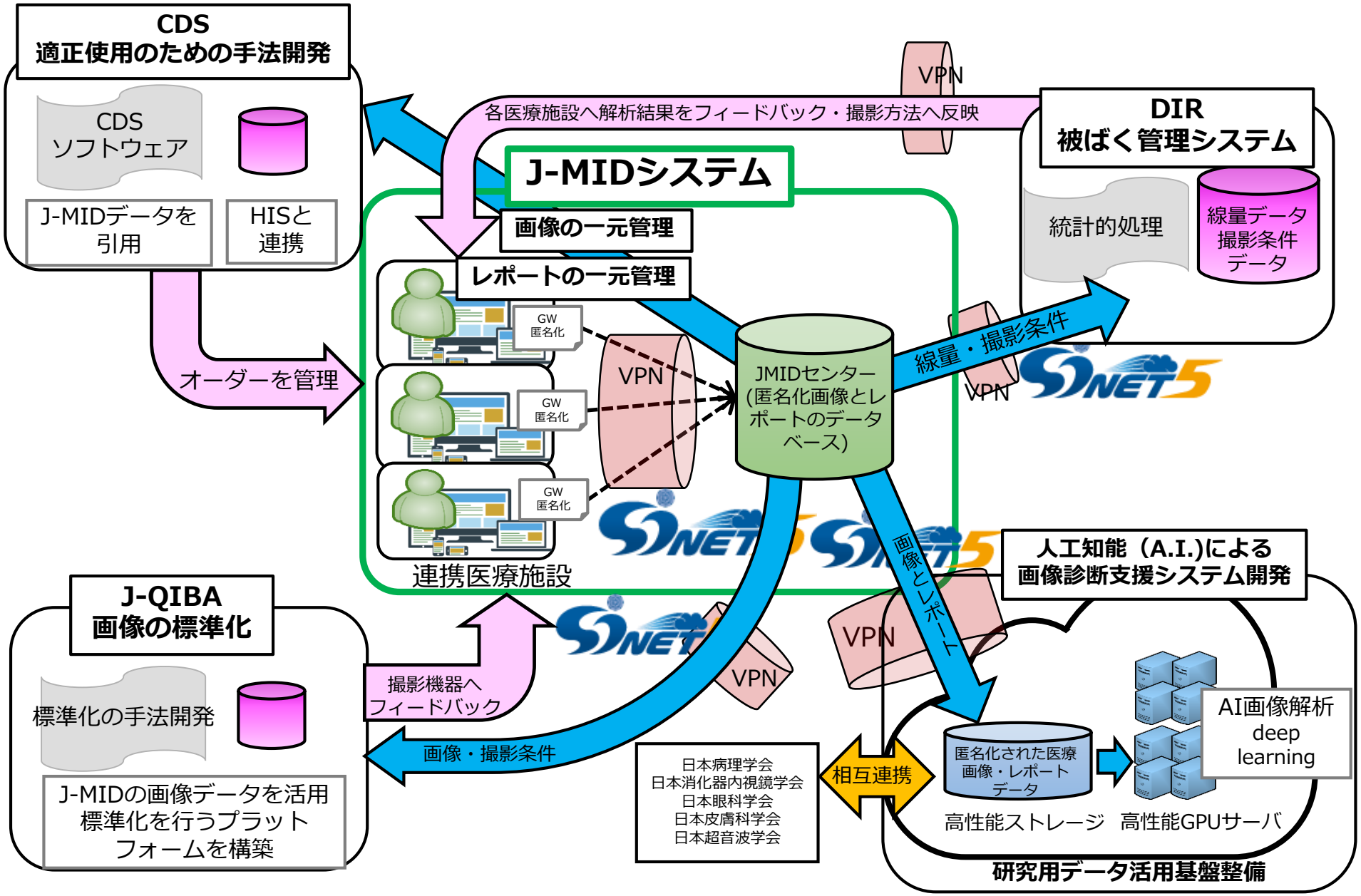
- 学術情報ネットワーク（SINET： Science Information NETwork）は、日本全国の大学、研究機関等の学術情報基盤として、国立情報学研究所（NII）構築、運用している情報通信ネットワーク。
- 教育・研究に携わる数多くの人々のコミュニティ形成を支援し、多岐にわたる学術情報の流通促進を図るため、全国にノード（ネットワークの接続拠点）を設置し、大学、研究機関等に対して先進的なネットワークを提供。
- 2016年よりSINET5の本格運用開始、全国100Gネットワークにて接続。

※右図回線構成図とともに、NII Webページより。

組織体制①: 日本病理学会（JP-AID）

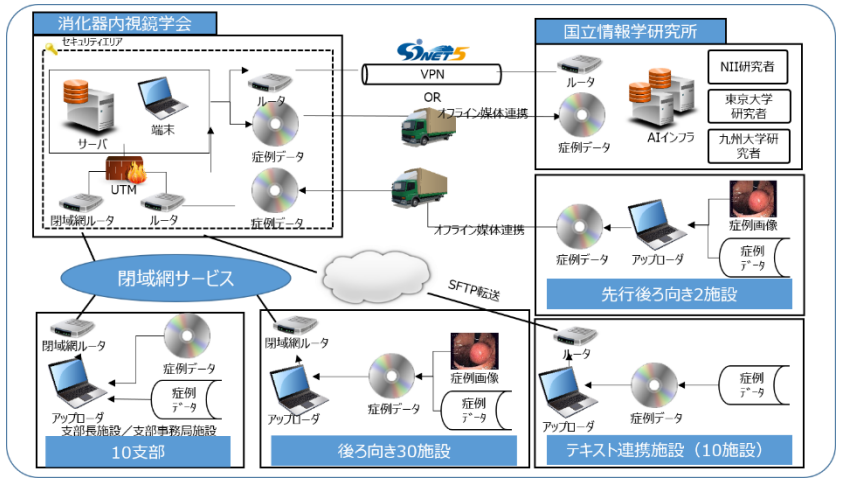


組織体制②：日本医学放射線学会（J-MID）



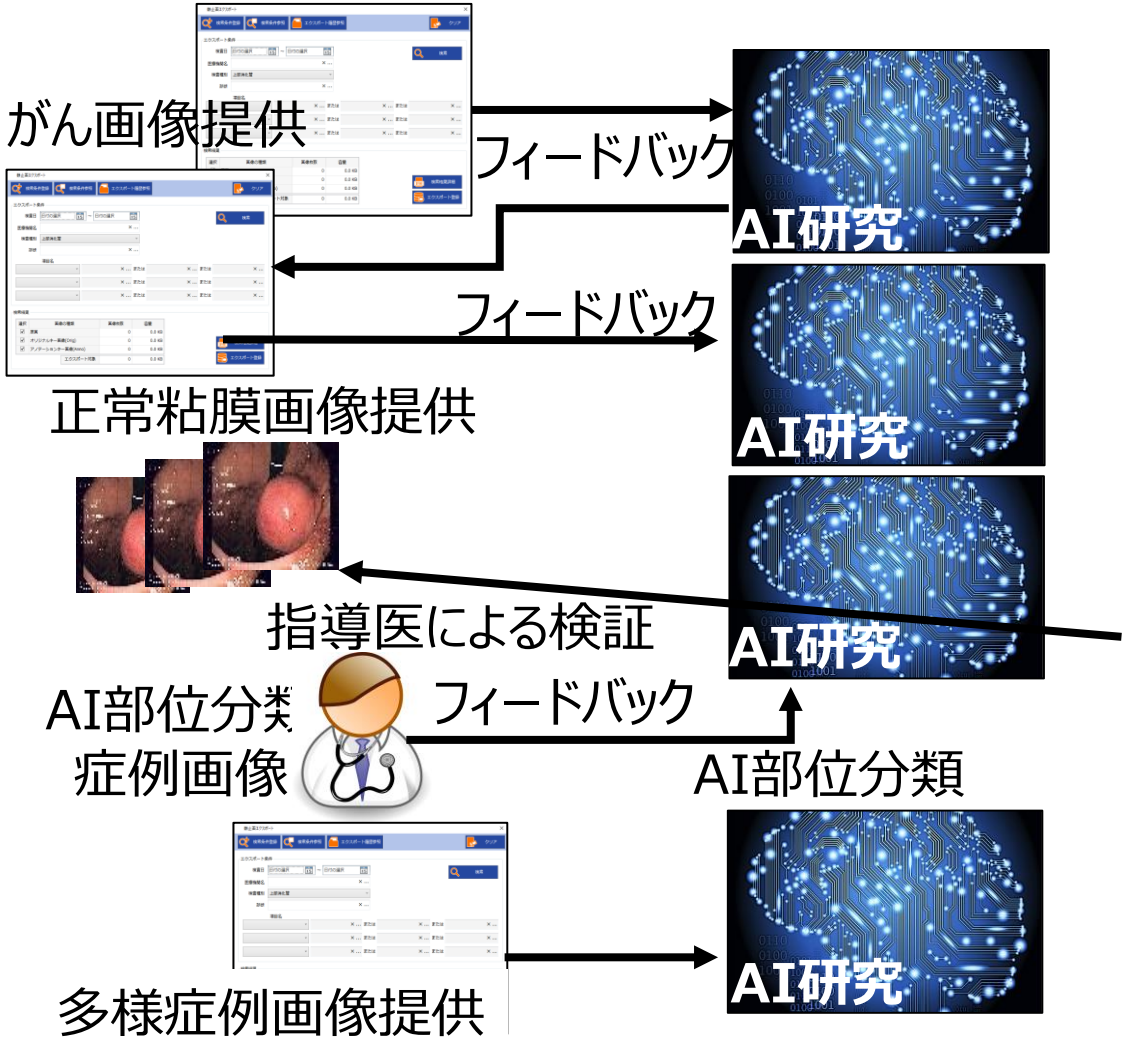
組織体制③：日本消化器内視鏡学会

情報収集基盤



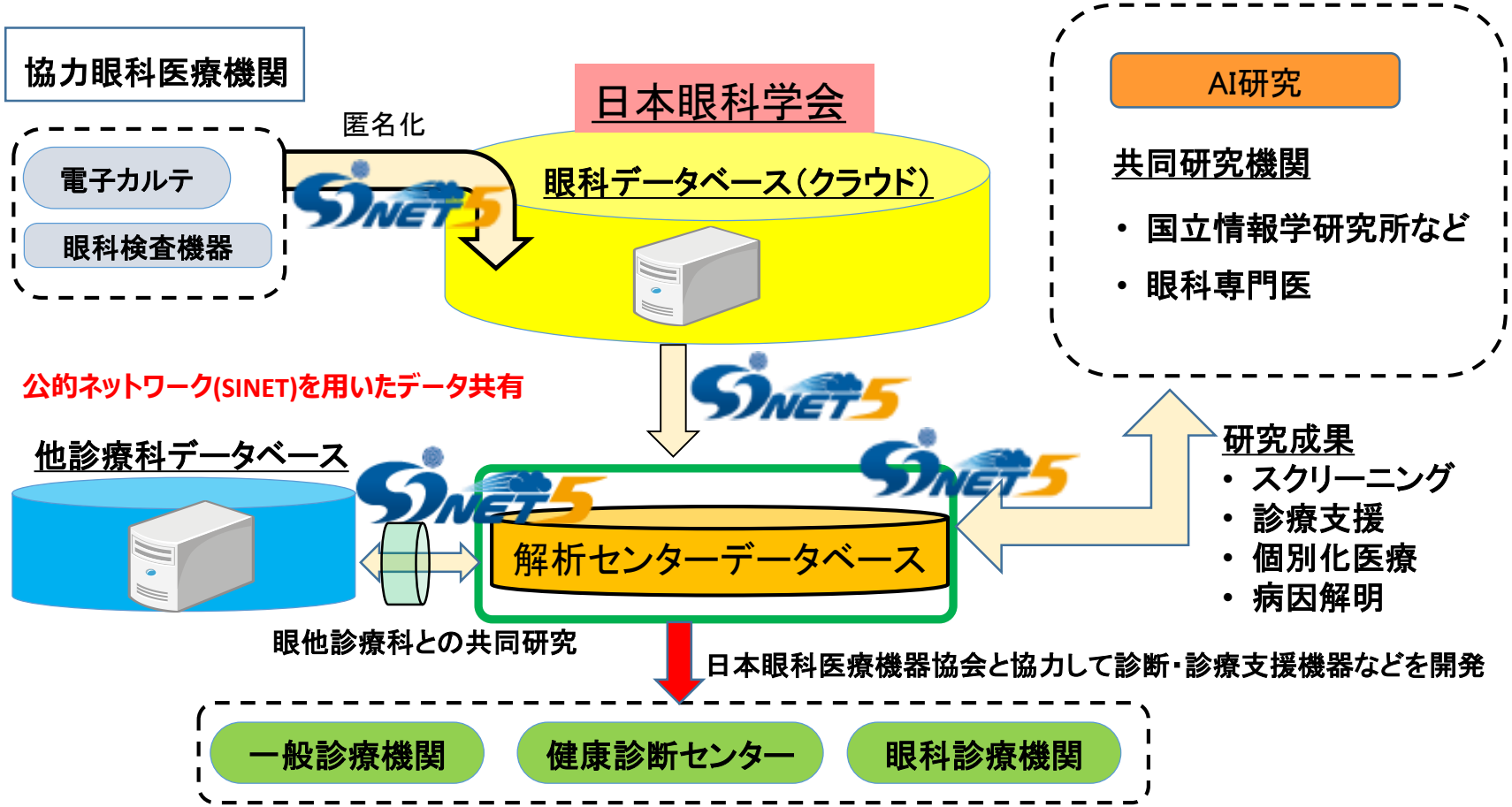
画像集積結果(2018/03まで)	
総合計枚数	1,728,424

臨床サイドとAI研究者とのインタラクティブな議論



組織体制④：日本眼科学会

日本眼科学会（学）、データサイエンティスト（学・産）、日本眼科医療機器協会（産）、AMED（官）

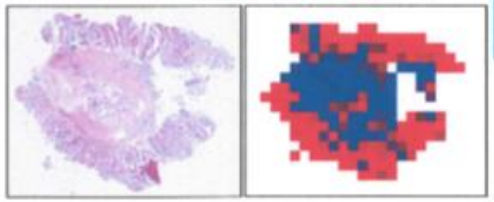


高いセキュリティーを確保したデータ管理体制内で継続的にデータ収集・解析を行い
常に新しい成果を挙げ、継続的にこれを医療・社会へ還元する

日本病理学会の成果

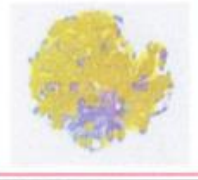
996症例の胃生検を機械学習させAIを開発

AIによる判定の例(赤：癌の可能性高い)

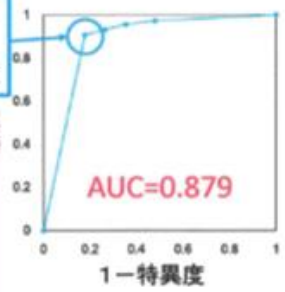


カットオフ値0.8の場合
病理医とAI判定の
不一致率16.2%を達成

重ね合わせ表示も可能



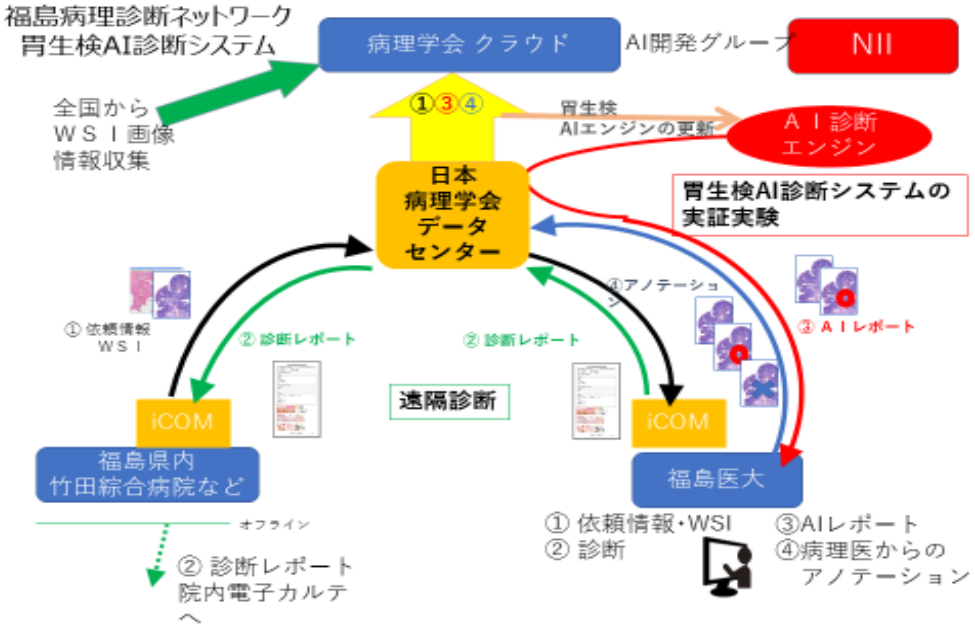
ROC曲線



Cut off 異常度0.6の場合

		AI判定 Normal	AI判定 Cancer
病理医の診断	腫瘍でない	1602 (73.5%)	579 (26.5%)
	癌・腫瘍(疑い含む)	34 (6.7%)	470 (93.3%)

AIによる胃生検ダブルチェック支援システム



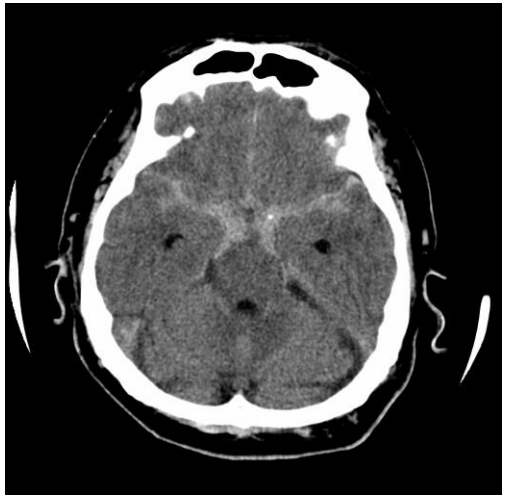
※WSI：Whole Slide Image
病理デジタル画像

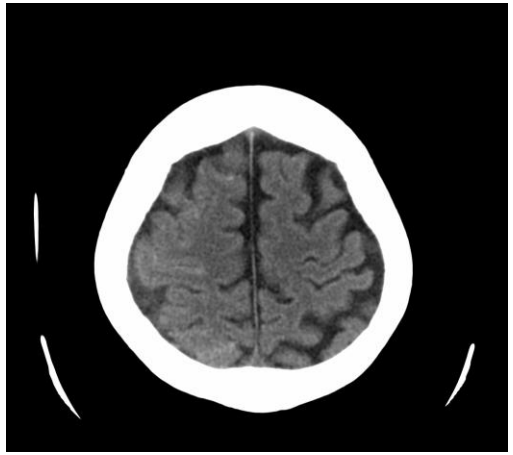
日本医学放射線学会の成果

2018年度のタスク
・頭部CTによるくも膜下出血(SAH)の検出

・第一段階 (データ提供済: 2018年9月初頭)
症例数: SAH 30例 + 非SAH 33例
スライス数: SAH有 602枚 + SAH無 1516枚
ラベル: 1スライスごとの「SAHの有無」

・第二段階以降
第一段階での感触をフィードバックして症例数の拡充
ラベル⇒アノテーションデータ



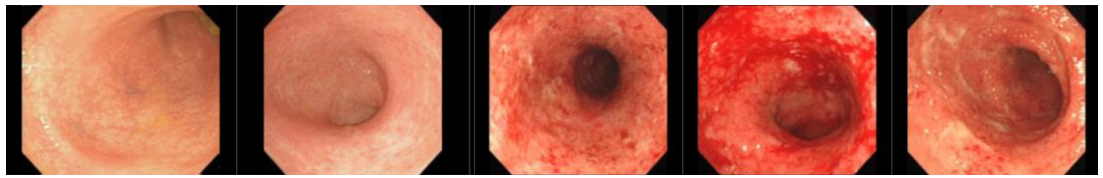


※AI画像解析担当：東京大学

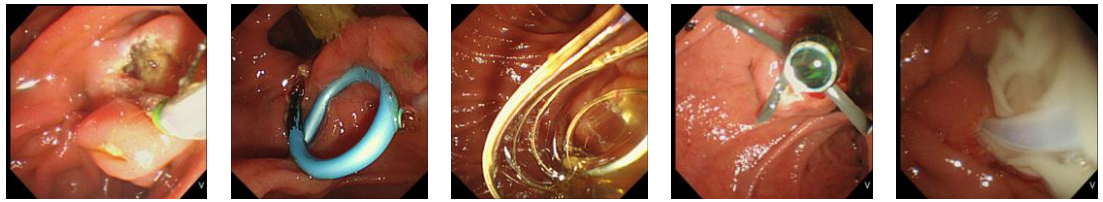
日本消化器内視鏡学会の成果

2018年度のタスク
・胃がん検出の精度向上：方法を学会で検討中
・潰瘍性大腸炎：炎症の程度や範囲で鑑別診断支援
・十二指腸乳頭：ERCPの偶発症予測

- 潰瘍性大腸炎
  - ・20,042 画像（画素数は一定ではない）
  - ・多様な部位と多様な症状がミックスされたデータセット
  - ・ラベルなし



- 十二指腸乳頭
  - ・17,369 画像
  - ・多様な状況がミックスされたデータセット
  - ・ラベルなし



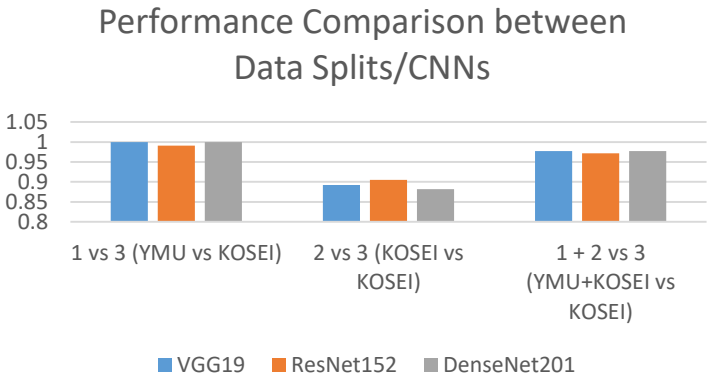
※AI画像解析担当：（部位認識）九州大学、（胃がん検出）東京大学

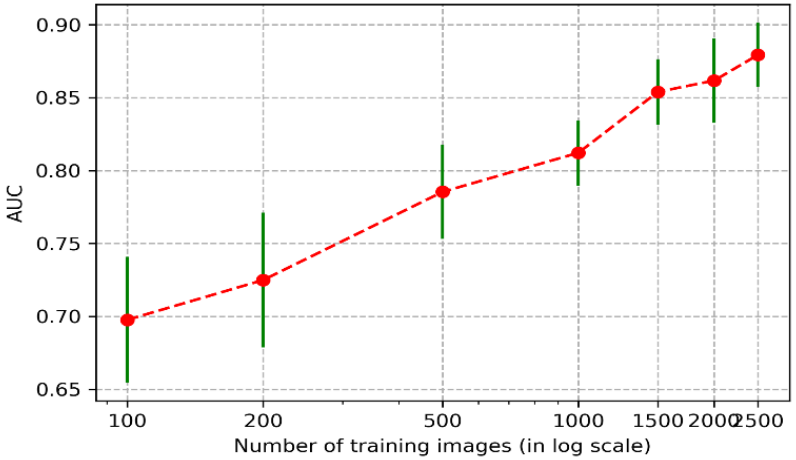
日本眼科学会の成果 ※一部追記

個々の患者さんでデジタルデータを経時的に追跡可能

2018年度のタスク
・緑内障の眼底画像の検出（昨年度からの継続）
・重症度情報のある疾患（緑内障、加齢黄斑変性症、糖尿病網膜症）の重症度分類の検討

緑内障
positive/negative

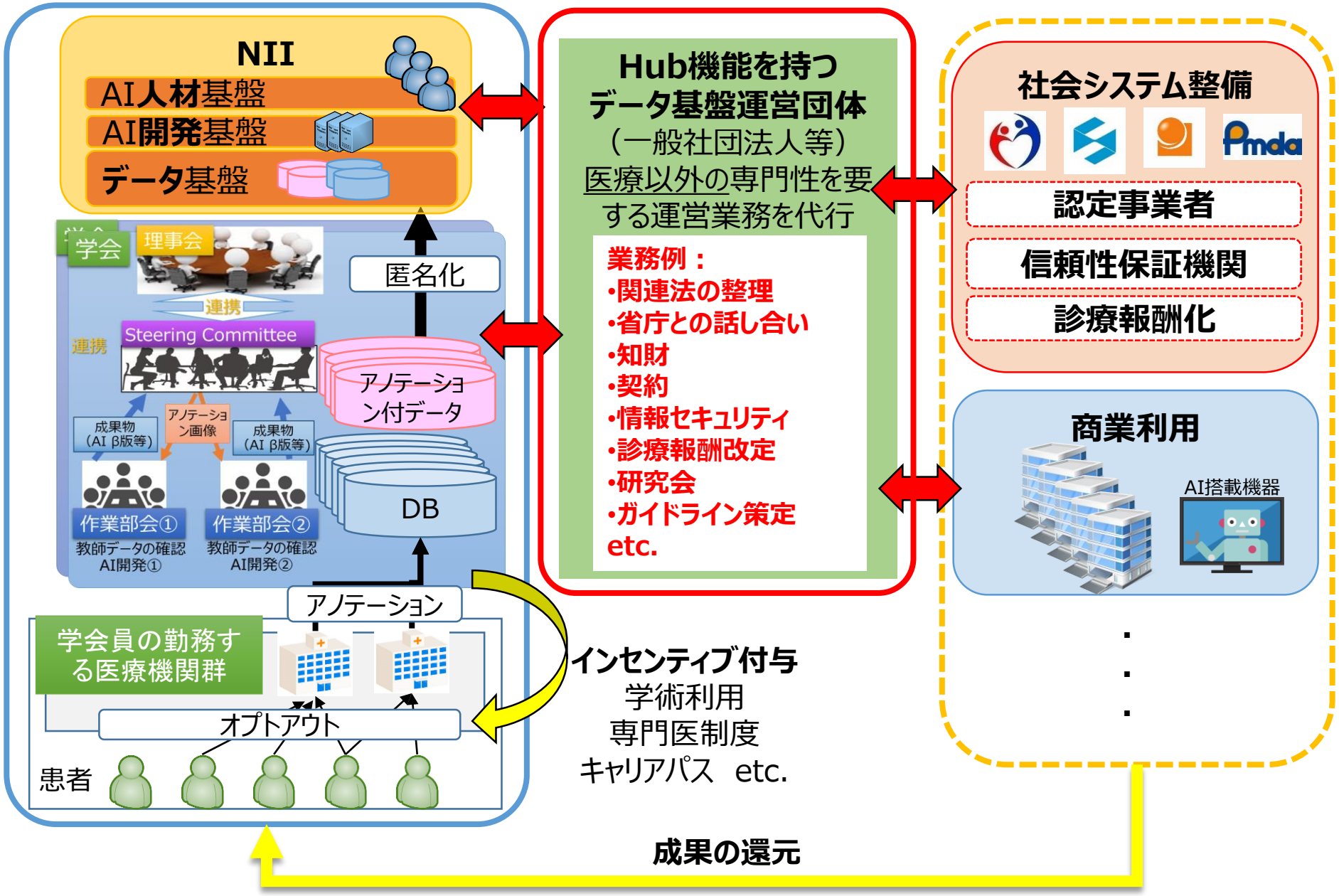




⇒更なる症例追加による
精度向上が期待される

※AI画像解析担当：NII

課題解決のための方向性（案）



課題・障壁：研究で解決できる部分と難しい部分

	課題・障壁		AMED研究の成果（計画含む）		残っている課題 （研究では解決困難）
Ⅰ．教師データセット作成に係る障壁	1.アノテーション ・正確性の担保 ・作業負担 ・必要な粒度が異なる	→ → →	・WGの構成（分野専門知識） ・ツール（アノテーションAI） ・NIIと協働（AI専門知識）	→ → →	（ー） ・制度・診療報酬化 （ー）
	2. 匿名化 ・「完全匿名化」の定義が曖昧 ・教師データの質が劣化	→ →	・（個人情報削除で対応） ・（上記と同様の対応）	→ →	・法・指針解釈の確定 （ー）
Ⅱ．データの第三者提供に係る障壁	3. 個人情報・倫理審査 ・オプトアウトの拡大解釈 ・企業等へのデータ提供不可	→ →	・（倫理審査委員会での説明） ・（個別同意を再取得？）	→ →	・法・指針解釈の確定 ・法・指針の改訂
	4. データセキュリティ ・情報管理者の負担増 ・ガイドライン等の不在	→ →	（ー） ・JAMIによるガイドライン	→ →	・制度・診療報酬化 （ー）
	5. データ標準化 ・データフォーマット、規格格納形式、HISからの抽出方法 ・付帯情報・読影レポート	→ →	・（学会内である程度統一） ・学会内である程度統一	→ →	・標準規格の設定 （ー）
	6. 知的財産権 ・学会所有の画像データ ・教師データセットの所有権 ・AIアルゴリズム	→ → →	・（学会の所有物） ・（所有権が不明瞭） ・（知財の管理方法は？）	→ → →	・指針の策定 ・制度・診療報酬化 ・PMDA承認　など

参考資料

機密２

資料3

AIによる画像診断支援に向けた研究の進捗状況について

国立研究開発法人
日本医療研究開発機構（AMED）
理事長　末松　誠

平成30年7月23日
第1回保健医療分野AI開発加速コンソーシアム
於：厚生労働省

データ共有に係る課題について

今後学会の収集データは、広く公平性を持って提供していくことが望まれる。

しかし、
以下の問題が未解決である。

I.教師データセット作成に係る障壁

１．メタデータ付与（アノテーション）
２．匿名化

II.データの第三者提供に係る障壁

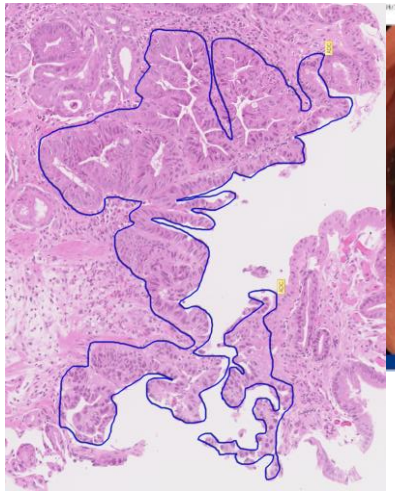
１．個人情報・倫理審査
２．データセキュリティ
３．データ標準化
４．知的財産権

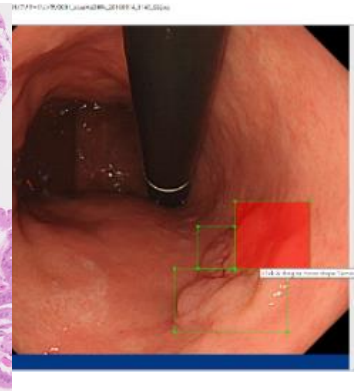
教師データセット作成に係る障壁

メタデータ付与（アノテーション）について

◆ 診断名付与について

- 診断名の正確性の担保はどの領域においても困難である。
- 臨床現場で確定診断名が付与できない領域（消化器内視鏡、放射線、超音波等）がある。





◆ 病変部位マーキングについて

- 臨床現場の通常業務ではない。
  （行っていてもAI用のアノテーション作業と異なる）
  →現場医師に追加の作業負担が生じる。
  例）病理画像：専門医の手で1枚数時間
  消化器内視鏡画像：熟練者の手で1枚2-3分程度
- 作成するAIの目的によって必要とされる粒度が異なる。

学会ごとにより以下の取り組みが行われている。

- アノテーション補助ツールの開発
- 学会内におけるアノテーションの為のWGの作成
- AIの目的ごとのメタデータの粒度につきNIIと相談

教師データセット作成に係る障壁

匿名化について

「人を対象とする医学系研究に関する倫理指針ガイダンス」によれば、適切なマスキングがされていれば良いと解釈できる。

第5章第12(3)ア(ウ)の「匿名化されているもの」とは、既存資料・情報の中に含まれる記述等のうち、その記述単体で特定の研究対象者を直ちに判別できる記述等を全て削除するような加工がなされているものを指す。ここでいう「記述単体での特定の研究対象者を直ちに判別できる記述等」とは、具体的には例えば、氏名、顔画像、個人識別符号（ゲノムデータ、保険証番号等）が該当する。病名、検査データ等については、その記述等が比較的特異な場合であっても、基本的には「その記述単体で特定の研究者を直ちに判別できる記述等」には該当しないものと取り扱ってよい。

しかし、

- 一部の画像については依然「完全匿名化」の定義が曖昧である。（再構築可能データと個人の認識が可能なデータとの混乱）
- 完全匿名化により教師データとしての質が劣化する可能性がある。

データの第三者提供に係る障壁

個人情報・倫理審査について

現在は、「インフォームドコンセントの手続き等の簡略化（医学系指針第12の７）に該当するとしてオプトアウトで画像を収集している。

７ インフォームド・コンセントの手続等の簡略化

(1) 研究者等又は既存試料・情報の提供を行う者は、次に掲げる要件の全てに該当する研究を実施しようとする場合には、研究機関の長の許可を受けた研究計画書に定めるところにより、１及び２の規定※による手続の一部を簡略化することができる。

① 研究の実施に侵襲（軽微な侵襲を除く。）を伴わないこと。

② １及び２の規定※による手続を簡略化することが、研究対象者の不利益とならないこと。

③ 1及び２の規定※による手続を簡略化しなければ、研究の実施が困難であり、又は研究の価値を著しく損ねること。

④ 社会的に重要性が高い研究と認められるものであること。

※１及び2の規定：同章「インフォームド・コンセントを受ける手続き等」・「研究計画書の変更について」を指す。

しかし、

- 医療情報部、倫理審査委員会等の理解を得られないことがある。
  - 本当に実施困難か、社会的に重要性が高いか
  - 解釈が困難な中での拡大解釈
- 企業等へのデータ提供を想定した場合、現行方式では成り立たない

データの第三者提供に係る障壁

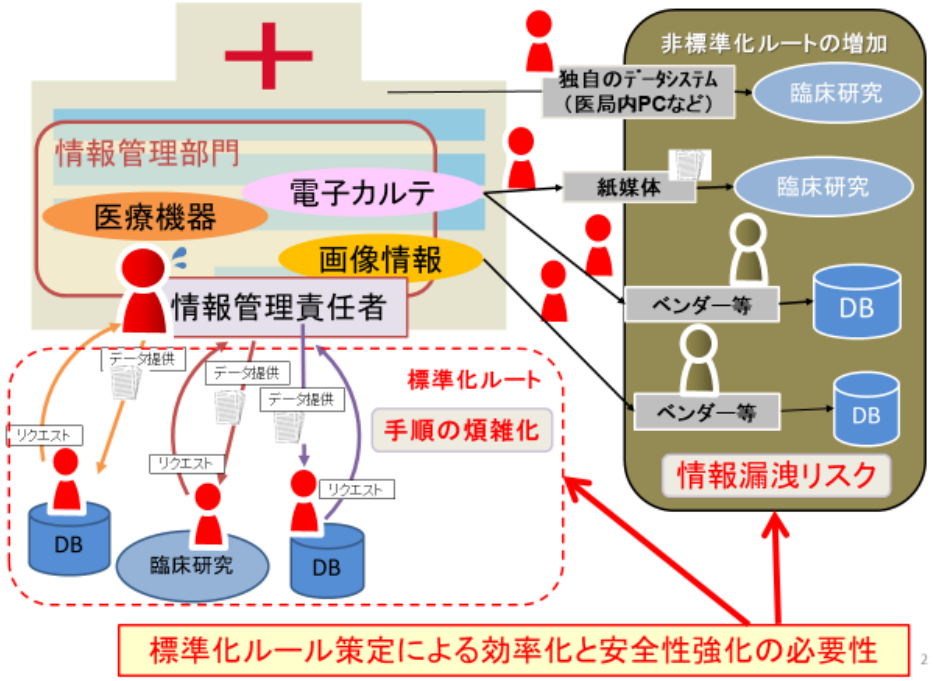
データセキュリティについて

倫理審査とは別に、情報管理者による情報漏えいの監督も必要である。

しかし、

- 情報管理者の負担増
- ガイドライン等の不在

により、説明・同意の取得が困難な場合がある。



AMEDでは本年度より、医療情報学会主導の研究開発を委託

データの第三者提供に係る障壁

データの標準化について

データ共有のためには、少なくとも以下においてベンダーニュートラルな標準化が必要である。

- データフォーマット
- 規格
- 格納形式
- HIS※からの抽出方法
- 付帯情報（項目・用語・書式）
- 読影レポートのフォーマットや用語

しかし、

- 全てを統一するのは困難である。
  （MRI画像のテスラ数、超音波画像の周波数など）
- 学会内である程度企画を統一しても、現場全体の統一は困難である。

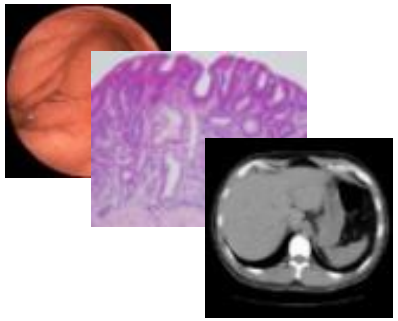
※HIS : Hospital Information System

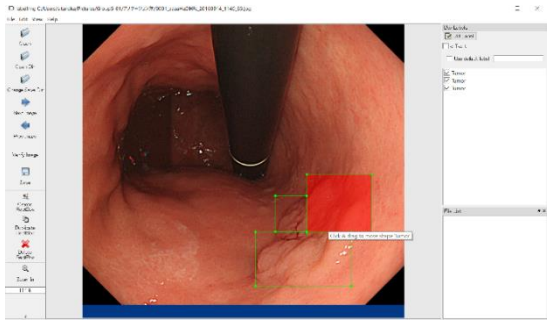
データの第三者提供に係る障壁

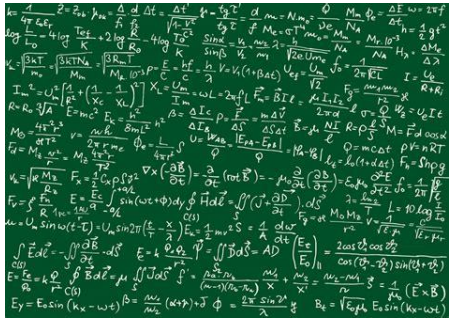
知的財産権について

広範かつ公正な共有が求められるデータには以下がある。

- 学会等の所有する画像データ
- 学会等の作成したアノテーション済み教師データセット
- 学会等が開発するアルゴリズム







しかし、

- データセットが無償提供されることにより、アノテーターのモチベーションが下がる可能性がある。
- 収集画像データは学会の所有物であるが、データセットの所有権については不明瞭である。
- ある研究者が類似したアルゴリズムの特許権を取ってしまうことで、研究に支障が生じる可能性がある。

課題解決に有用となり得る指針等（案）

◆倫理審査に係る課題について

- 医学研究・診療分野での医療情報の利活用の形態は多様化している。
- 医療情報を利活用した研究促進・社会還元を図るには、法的に容認される範囲、研究形態等の更なる明確化が望まれている。

◆標準化に係る課題について

- DICOM規格と同様に、画像データ全体に共通した収集・格納方法などの規格が望まれる。

◆第三者提供に係る課題について

- 同意取得や画像匿名化につき、法律上許容できる具体例の明示が望まれる。
- 医療等AIのデータの所有や知財の在り方（アノテーション済み教師データの所有権、AIの特許や知的財産権、知的財産の管理方法、研究段階での留意事項など）についての指針があることが望ましい。

◆公益性に係る課題について

- 優先的な開発を推奨するAI開発分野につき、国の指針を策定することが望ましい。