

データ利活用プラットフォームの 提供サービス（CANNDs）について

2021年10月20日

国立研究開発法人 日本医療研究開発機構 (AMED)

1. CANNDsの定義

「CANNDs*（キャンディーズ）」とは、AMEDが支援した研究開発から得られたデータの利活用を促進するためのプラットフォーム（AMEDデータ利活用プラットフォーム）であり、総合的なサービスの総称である。

2. 概要(コンセプト)

- ✓ AMED研究で生み出される研究開発データの利活用を促進するサービスを提供する。
- ✓ 必要とされる研究開発データへアプローチする利便性を高め、AMED研究開発データ利活用カタログにリストアップされた提供データを利活用する際の手続きを簡略化する。
- ✓ 機微なデータを持ち出すことなく様々な検索・解析ができるよう、セキュリティを強固にした閉鎖型の計算環境を構築することで、データ提供者とデータ利用者の双方が安心して、データを利活用できる環境を提供する。

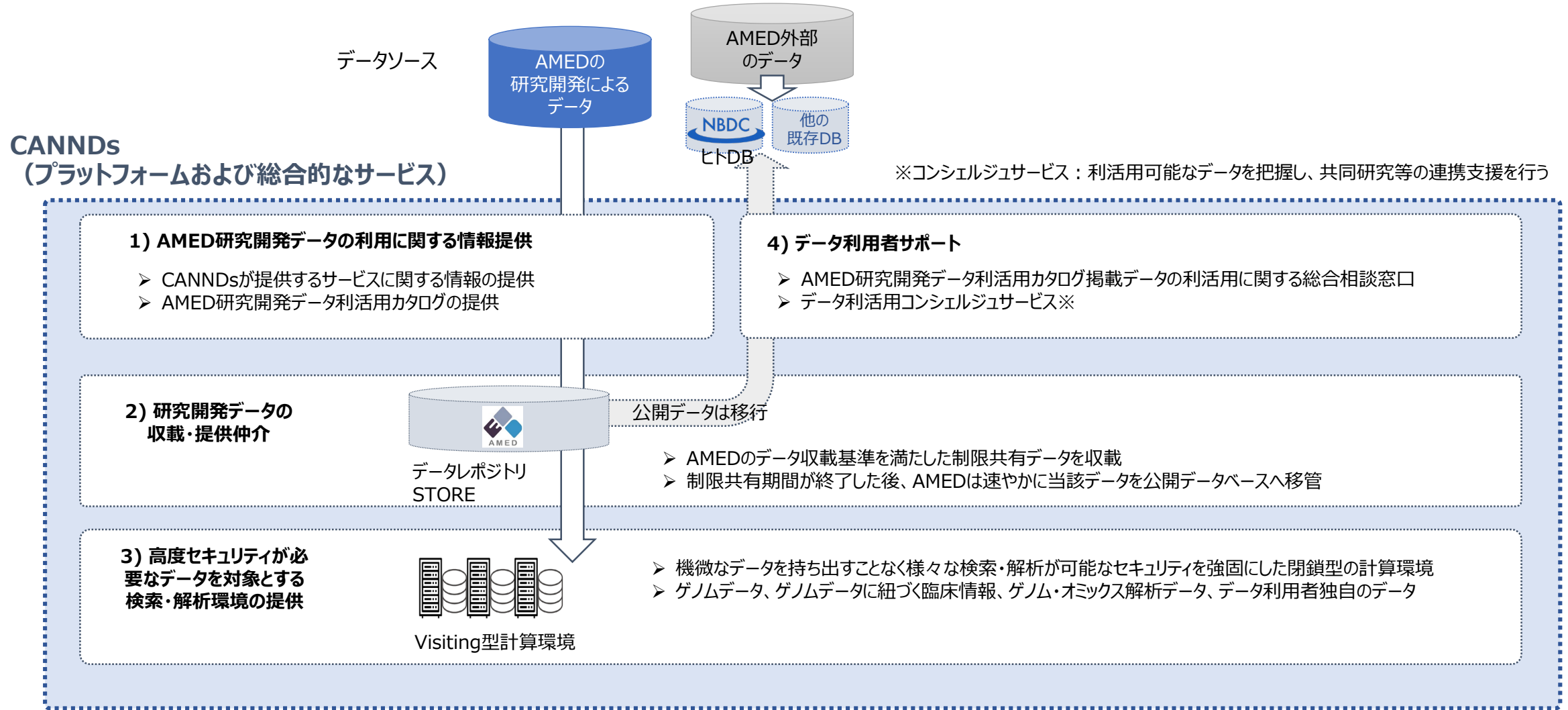
3. 主な提供サービス

CANNDsで提供するサービスは以下の通りである。

- 1) AMED研究開発データの利用に関する情報提供
- 2) 利活用が可能な研究開発データの収載・提供仲介
- 3) 高度セキュリティが必要なデータを対象とする 検索・解析環境の提供
- 4) データ利用者サポート

AMEDが支援した研究開発により得られたデータの利活用プラットフォーム

既存のデータベース等と連携・役割分担しつつ、AMED研究で生み出される研究開発データの利活用を促進するサービスを提供する。



★データの共有の考え方 (AMED 研究データ利活用に係るガイドライン1.1版より)

非制限公開：データマネジメントプランに基づいてデータベースに登録することにより、当該データベースの規約等に従う限り、アクセスに制限なく誰でも利用することが可能となるデータ。

制限公開：データマネジメントプランに基づいてデータベースに登録することにより、当該データベースの規約等に従って、利用目的、利用方法を明らかにしたうえで、データアクセス申請を承認された研究者等が利用することが可能となるデータ。

制限共有：データマネジメントプランに記載された研究者等又は AMED が個別に承認した研究者等のみが、データベースの閲覧等又はその他の方法により利用することが可能となるデータ。

【年度別仕様（Visiting型計算環境）】

	R4年度	R5年度
解析環境	Visiting型計算環境下に設置した利用者専用アカウント	Visiting型計算環境下に設置した利用者専用アカウント
利用できるデータ	コントロール群のゲノムデータ（FASTQ, BAM/CRAM,gVCF,VCF）およびメタデータ（年齢、性別、出生地／居住地、疾患名）	がん・難病のゲノムデータ、それに紐づく臨床情報および症例のメタデータ コントロール群のゲノムデータ（同）およびメタデータ（同）（調整中）
	利用者が持ち込んだデータ （制限公開・非制限公開データを含む）	ゲノム・オミックスデータ 利用者が持ち込んだデータ（同）
プレリサーチ※5	無	有
利用できるソフトウェア	各種2次解析用オープンソースソフトウェア および統計解析ソフトウェア（R）など	+データ利用者要望の元にソフトウェアなど追加検討
スパコン等計算機資源	TMM, NIG, IMS, NCC	TMM, NIG, IMS, NCC
ストレージ	TMM, NIG, IMS, NCC, NCGM	TMM, NIG, IMS, NCC, NCGM
利用方法	遠隔セキュリティルーム※1 リモートデスクトップ※2	遠隔セキュリティルーム リモートデスクトップ
データインポート	可（但し、審査承認が必要）	可（但し、審査承認が必要）
データエクスポート	個人情報を含まないもののみ可（但し、審査承認が必要）	個人情報を含まないもののみ可（但し、審査承認が必要）
独自ソフトウェアの追加・利用	可（但し、審査承認が必要）	可（但し、審査承認が必要）
メタデータ	サマリーシート※3	検索システム※4
利用料金	スパコン利用料のみ	+データ利用料（がん・難病）
利用対象者	利用登録をした一般の利用者 AMEDが指定する解析を行う利用者	利用登録をした一般の利用者

※1 遠隔セキュリティルーム：生体認証で厳重に入室管理されたCANNDs専用の解析ルーム。そこに設置したシンクライアント端末を通じ、CANNDsのVisiting型計算環境内に割り当てられたデータ利用者専用区画へのアクセスが可能。AMED施設内他、研究拠点に設置。

※2 リモートデスクトップ：CANNDsのVisiting型計算環境をデータ利用者のパソコンから操作するための仕組みです。データ利用者の所属機関からCANNDsの用意する連携スパコンを遠隔操作できるようになり、遠隔セキュリティルーム内のシンクライアント端末と同じ画面上で仕事ができる。

※3 サマリーシート：CANNDsにて利用可能なゲノムデータに対して、一次解析して得られた統計情報をサマリーシートとして開示。データ利用者は、サマリーシートをもとに、対象となるゲノムデータコレクションの利活用の判断に役立てることが可能。

※4 検索システム：ゲノムデータに紐づく属性情報（メタデータ）を使って解析対象となるデータを絞り込む機能を有するシステム。

※5 プレリサーチ：Genomics Englandで提供しているサービスを参考とし、研究計画書の作成前に、研究計画の検討を目的として、ゲノム情報と臨床情報の閲覧と検索ができるサービス。