

ゲノム医療実現のためのデータシェアリングポリシー

I. 目的

国立研究開発法人日本医療研究開発機構（以下「AMED」という。）が実施する医療研究開発において、正確な臨床・健診情報が付加されたゲノム情報を、データシェアリングにより利活用し研究を推進することが、国民の健康を向上させ疾患を克服するために重要であると考えている。

「第5期科学技術基本計画」（平成28年1月22日閣議決定）は、オープンサイエンスの推進により、学界、産業界、市民等あらゆるユーザーが研究成果を広く利用することで、研究者の所属機関、専門分野、国境を越えた協働による知の創出を加速し、新たな価値を生み出していくことを目指している。また、研究分野によって研究データの保存と共有の方法に違いがあることを認識し、国益等を意識したオープン・アンド・クローズ戦略及び知的財産の確保等に留意することが重要であると言及している。

ゲノム医療実現推進協議会「中間とりまとめ」（平成27年7月健康・医療戦略推進本部）も、研究の推進のため、正確な臨床・健診情報が付加されたゲノム情報等のプロジェクト間でのデータシェアリングが重要であると指摘している。

データシェアリングは、データ計測・取得の重複実施を回避する等の効率性の観点と、他者のデータとの分析により重要且つ新たな洞察を取得する等の効果性の観点の両面から、推進していかなければならない。しかし、研究の企画や遂行に貢献することなく、一方的にデータを得て研究を行う研究者等（いわゆる **research parasite**）¹ が出現し、苦勞してデータを取得・分析し共有・公開した研究者のモチベーションを著しく阻害するおそれも考えなければならぬ。さらに、検体や臨床情報を提供した研究参加者の権利保護についても、関係法令、倫理指針等に基づく個人情報の保護を含め十分な対応が必要である。

こうしたことから AMED では、「疾患克服に向けたゲノム医療実現プロジェクト」、「脳とこころの健康大国実現プロジェクト」、「新興・再興感染症制御プロジェクト」、「難病克服プロジェクト」及び肝炎等克服実用化研究事業、エイズ対策実用化研究事業が資金提供を行う研究事業を対象として、以下のとおりデータシェアリングポリシーを定め、本ポリシーの適用となる研究開発課題については、原則としてデータシェアリングを義務づけることとした。本ポリシー²は、ゲノム情報を用いた医療の実現に向け、研究成果に紐づくゲノムデータ、及び臨床情報や解析・解釈結果等を含めたゲノム情報の、迅速、広範かつ適切な共有・公開を行うことを目的として、研究の対象者として参加する患者や健常者等の権利保護、データ・情報を提供した研究者の権利保護と、データシェアリングによる関連分野の研究の推進を両立するための枠組みを示すものである。

II. 用語の定義とデータシェアリング（データの共有・公開）の分類

1. 本ポリシーにおいて「ゲノム情報」とは、ヒト又はヒトに影響を与える微生物等を対象

¹ Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D., “Data Sharing”, *New England Journal of Medicine*, Vol. 374, No.3, 276-277 (2016)

² 臨床情報など、ゲノム情報以外のデータシェアリングポリシーについては今後別途策定する予定。

としたゲノム解析 (GWAS、SNP アレイ、ゲノムシーケンス、トランスクリプトーム解析、メタゲノム解析、エピゲノム解析、遺伝子発現解析等) によって得られた情報及びそれに関連する情報のうち、以下に規定するものをいう。

- (1) 生殖細胞系列、体細胞由来 DNA から得られる塩基配列情報
- (2) 生殖細胞系列由来、体細胞由来 DNA 等に存在する多型情報・変異情報
- (3) 後天的に生じるゲノム変化 (がん細胞等に生じた体細胞変異)
- (4) 遺伝子発現プロファイル、ゲノム修飾等
- (5) 健康に影響を与え得る微生物群等由来 DNA 等の塩基配列情報
- (6) 健康に影響を与え得る微生物等由来 DNA 等に存在する多型情報・変異情報
- (7) 関連する表現型情報・臨床情報のうち、AMED が指定する情報
- (8) その他 AMED が指定する情報

2. 本ポリシーにおいて「データマネジメントプラン」とは、研究開発課題ごとに研究開発代表者が定める、ゲノム情報の保存・共有・公開方針を記載した文書であり、IV. 2. に規定するものをいう。

3. 本ポリシーにおいて「データベース」とは、ゲノム情報及びそれに付随する情報を記録するためのデータストレージ (自機関や代表機関等に保管 (データマネジメントプランに記載することにより AMED に対し報告) するものをいう。)、AMED Genome group sharing Database (AGD)³、Medical Genomics Japan Variant Database (MGeND)、難病プラットフォーム、その他 AMED が指定する公的データベース (NBDC ヒトデータベース (JGA/NHA/DRA 等) 等) をいう。

4. 本ポリシーにおけるデータの共有・公開の範囲は次の3つに分類され、データの登録及び共有・公開の方法は「データマネジメントプラン」に基づくこととする。

- (1) 制限共有データ：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、データマネジメントプランに記載された研究者、及びデータアクセス申請を承認された研究者間で共有するデータ⁴。データの共有は原則的に研究者間の合意に基づき行うこととし、対象となる研究者は以下のとおりとする。なお、必要に応じて AMED が調整を行うことがある。
 - 当該研究グループの既存データの拡充・充実等に資するデータを提供する、又は今後提供し得る研究者等
 - データ生産や品質向上、付加価値付け等に貢献・協力できる研究者等
 - データの蓄積、活用等に貢献・協力を期待できる研究者等
 - その他、AMED が指定する者

³ AGD は、制限共有データを効率的・効果的に運用するため、NBDC (国立研究開発法人科学技術振興機構バイオサイエンスデータベースセンター) と DDBJ (国立遺伝学研究所 DNA Data Bank of Japan) の協力の下、AMED が構築した公的データベースであり、その運用費用は、AMED が負担している。

⁴ データそのものは非公開であるが、データ共有の実施状況は一部公開することができる (IV. 3. (1) 参照)。

- (2) 制限公開データ：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、利用方法等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が利用することが可能なデータ。
- (3) 非制限公開データ：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、アクセスに制限なく誰でも利用することが可能なデータ。

III. 本ポリシーを適用する研究開発課題の範囲

- 1. 本ポリシーは、平成31年度以降に開始する以下の(1)かつ(2)に該当する研究開発課題に対し適用する。ただし、本ポリシーの改定があった場合において、改定前に本ポリシーの適用を受けて開始している研究開発課題については、改定前のポリシーを適用する。
 - (1) AMEDが資金提供を行う研究開発課題のうち、ゲノム情報を生成するもの。
 - (2) AMED公募要領にデータシェアリングポリシーを適用することを明記しているもの。

IV. 「データシェアリングの実施方法」及び「データマネジメントプランの作成」等

1. データシェアリングの実施方法

- (1) 制限共有データの対象となるデータ等⁵については、原則として、「生データ生成後2年」又は「研究成果の公表時（論文採択、特許出願公開等の日をいう。以下同じ。）」のいずれか早い時点までに、制限共有データの登録先としてAGDに登録、又は自機関や代表機関等に保管し⁶、共有することとする。なお、倫理的配慮や商業的機密情報であること、社会に危険をおよぼす微生物のゲノム情報であること等により制限共有を実施することが困難な場合は、この限りではない。
- (2) 制限公開データの対象となるデータ等⁷については、原則として、「生データ生成後2年」又は「研究成果の公表時」のいずれか早い時点までに、制限公開データの登録先としてAMEDが指定する制限公開データベース⁸へ登録し公開することとする。なお、倫理的配慮や商業的機密情報であること、社会に危険をおよぼす微生物のゲノム情報であること等により制限公開を実施することが困難な場合は、この限りではない。
- (3) 非制限公開データの対象となるデータ等⁹については、原則として、「生データ生成後2年」又は「研究成果の公表時」のいずれか早い時点までに、非制限公開データの登録先としてAMEDが指定する非制限公開データベース¹⁰に登録し公開することとする。AMEDが指定する公的データベース以外を非制限公開データの登録先とする場合はAMEDと協議する。

⁵ 個別のBAM及びVCF等のデータ、及び個人毎の臨床情報が入った疾患データベース等のデータ等を想定。

⁶ 自機関や代表機関で保管する場合も、研究の対象者として参加する患者や健常者等に対して、本ポリシー規定に従い、当該研究者以外にデータが共有される可能性があることについて同意を取得しておくこと（V.2.参照）。

⁷ 個別のBAM及びVCF等のデータ等を想定。

⁸ NBDCヒトデータベース（JGA/NHA/DRA等）等の公的なデータベースのこと。

⁹ 個人の特定が困難な集団における統計データ等を想定。

¹⁰ NBDCヒトデータベース（JGA/NHA/DRA等）やMGeND等の公的なデータベースのこと。

2. データマネジメントプランの作成

(1) データマネジメントプラン (AMED が指定する様式¹¹) には、以下の事項について記載する。ただし、一部の事項を省略することができる。

- 事業年度【必須】
- 事業名【必須】
- 研究開発代表者の氏名、所属【必須】
- 研究開発課題名【必須】
- データマネジメントプランの対象となる分担課題名
- 研究から産出されるデータ及びデータ群の総称【必須】
- 研究開発データの説明【必須】
- リポジトリ (データの保存場所)【必須】
- データサイエンティスト¹²の氏名、所属、研究者番号 ※複数名を記載可【必須】
- 研究開発データを取得又は収集した者
- 研究開発データの管理者
- 自主管理データ、委託者指定データの別
- 研究開発データの想定利活用用途
- 研究開発データの取得又は収集方法
- 研究開発データの利活用・提供方針
- (他者に提供する場合) 円滑な提供に向けた取組
(秘匿して自ら利活用する場合) 秘匿理由・期間
- 想定データ量
- 加工方針
- その他

(2) データマネジメントプランには、制限共有データ、制限公開データ、非制限公開データの各々につき、以下の事項を AMED が指定する様式¹³に記載し、添付する。

- 登録するデータベース (構築予定のデータベースを含む)
- 登録の時期
- 対象データの種類、規模
- 公開・共有の範囲¹⁴

(3) データマネジメントプランは、全体計画書及び毎年度の研究計画書に添付し、全体計画書及び研究計画書の一部として取り扱う。

(4) 公募事業の事前評価において、データマネジメントプランの記載内容を評価の対象とする事業については、公募要領の定めるところにより提案書にデータマネジメントプランを添付する。なお、必要に応じて、課題評価委員会等からの採択条件として、デー

¹¹ 事業毎に別途通知する様式。様式例は別紙1を参照。

¹² 本ポリシーにおいて「データサイエンティスト」とは、単にデータを集約して処理するだけでなく、データの質を担保するとともにデータから有用な知見を引き出す者であって、データマネジメントプランの実行に責任を有する研究者として研究開発代表者が指名した者をいう。ただし、研究開発代表者が兼務することを妨げない。

¹³ 事業毎に別途通知する様式。様式例は別紙2を参照。

¹⁴ 企業等の参加も想定し、公開・共有の範囲を設定するものとする。

タマネジメントプランの修正を求める場合がある。

3. データマネジメントプランの実施状況の把握と評価等

- (1) 研究開始後、研究開発代表者は、AMEDからの求めに応じ、データマネジメントプランに記載されたデータの登録・共有・公開の準備及び実施状況（変更等も含む）を指定様式¹³に記載し報告するものとする。AMEDは、報告内容を進捗状況の把握に活用することに加え、その概要を一部公開することができる。なお、研究開発期間終了後も、AMEDはその進捗状況を確認することができる。
- (2) 研究開発代表者はデータ等の取扱いに不適切な利用（第三者への提供、目的外使用等）があったと認めた場合、又はそのおそれがあると認めた場合は、データの適切な管理のために必要な措置を講じなければならない。制限共有データにアクセスできる者によるデータ等の取扱いにあって、不適切な利用が判明した場合には、研究開発代表者はただちに該当者のデータ利用を停止し、AMEDに報告しなければならない。
- (3) 公募事業の事前評価において、過去のデータマネジメントプランの実施状況の評価結果、既存の解析データの登録についても、評価対象とすることができる。
- (4) 中間・事後評価においては、データマネジメントプランの実施状況の評価項目として掲げた上で評価の対象とする。なお、この評価項目においては、産出されたデータの登録・共有・公開の状況の評価するものとし、データ産出の進捗については評価対象としない。

V. 個人情報保護及び倫理的配慮

1. 研究の実施にあたっては、研究の対象者等に対する十分なプライバシー保護に配慮しなければならない。また、インフォームドコンセントの取得にあたっては、データの共有・公開を通じて様々なゲノム研究を実施し得ることについて規定を盛り込まなければならない。
2. データの共有・公開の実施にあたっては、国の定める法令及び倫理指針等を遵守しなければならない¹⁵。

VI. 知的財産

AMEDは、医療分野の研究開発の推進及びその成果の円滑な実用化に向け、知的財産の的確な保護を奨励しており¹⁶、研究者はAMEDの資金提供によって得られたデータをもとに知的財産権を取得できる。ただし、二次的研究の実施や、それにより得られる成果の実用化の機会を増やすため、知的財産権によりデータ共有が過度に妨げられないよう配慮しなければならない。

¹⁵ 「ヒトゲノム研究に関する基本原則について」（平成12年6月14日 科学技術会議生命倫理委員会）、「ヒトゲノム・遺伝子解析研究に関する倫理指針」（平成13年3月29日 文部科学省、厚生労働省、経済産業省）、「NBDCヒトデータ共有ガイドライン ver.4.0」（平成30年8月31日 NBDC）、「NBDCヒトデータグループ共有ガイドライン ver.2.0」（平成30年8月31日 NBDC）等を参照。

¹⁶ AMED知的財産ポリシー（平成27年4月1日 平成27年規程第27号）を参照。

VII. その他

1. 本ポリシーは、法律や指針の改正等により変更する可能性がある。
2. 商業的利用、当事者間の契約等その他事項については、必要に応じて別途定める。
3. AMED は、AGD に登録されたデータの利用を促進するために必要な措置を講じることができる¹⁷。

【参考】データの利用に関して

- (1) データ利用者は、成果発表時に登録データの利用について、データの提供元となった研究論文の引用を行い、以下の例を参考とした謝辞を述べることとする。

****【謝辞の例】**

「本研究に使用したデータ（の一部）は、国立研究開発法人日本医療研究開発機構〇〇事業の研究開発課題「AAAA」（研究開発代表者 BBBB）によって取得され、科学技術振興機構（JST）の「バイオサイエンスデータベースセンター（NBDC）」ウェブサイト（<https://biosciencedbc.jp/>）を通じて提供されたものです。」

- (2) データ利用者は、非制限公開データ及び制限公開データの二次利用による研究成果をもとに知的財産権を取得できる。

改定履歴

平成 28 年 4 月 策 定

平成 29 年 10 月 改 定

- ・適用範囲を「難病克服プロジェクト」に拡大

平成 30 年 11 月 改 定

- ・適用範囲を「脳とこころの健康大国実現プロジェクト」、「新興・再興感染症制御プロジェクト」、肝炎等克服実用化研究事業、エイズ対策実用化研究事業に拡大
- ・データマネジメントプランの作成及び実施状況の把握のための様式例の追加

¹⁷ 例えば、データ提供者の了解を得た上で、AGD に実装されたアレル頻度を知るための検索ツールの検索対象とすること等を想定

(別紙1) データマネジメントプラン様式

(作成日: 年 月 日)

		AMED課題管理番号(※AMEDで記入します)
事業年度		
事業名		
研究開発課題名		
研究開発代表者	所属	
	役職	
	氏名	
研究開発から産出されるデータ及びデータ群の総称		
研究開発データの説明		
リポジトリ(保存場所)		
データサイエンティスト 1 (データ解析の高い能力を有し、かつ、AMEDの研究目的を達成するためにデータの収集、質の確保、意味づけ、保存と活用等を行う研究者) ※複数数の場合、適宜行を追加	所属	
	役職	
	氏名	
	所属・氏名の公表可否	
	e-Rad番号	
	人件費の財源	
	雇用形態	
データサイエンティスト 2 (データ解析の高い能力を有し、かつ、AMEDの研究目的を達成するためにデータの収集、質の確保、意味づけ、保存と活用等を行う研究者) ※複数数の場合、適宜行を追加	所属	
	役職	
	氏名	
	所属・氏名の公表可否	
	e-Rad番号	
	人件費の財源	
	雇用形態	
その他(各事業で指定する事項)		

※記載の内容に変更があった場合は、その都度、速やかに変更後のデータマネジメントプランを提出してください。

(別紙2) データマネジメントプランに記載されたデータの登録・共有・公開の準備及び実施状況(様式例)

記述 の 注 意	登録 の 注 意	研究開発から提出される データ及びデータ類のうち、 データシェアリングポリシー の対象となるデータの概要 ※注1 ※注2	対象者の 年齢 ※注2	対象者の性別 (1)~(6)の欄に 記載)	解析データの 解析方法 (データ分類 (1)~(6)の欄 に記載) ※注3	解析データの 解析方法 (データ分類 (1)~(6)の欄 に記載) ※注4	関連する表現 で示した データ分類 (データ分類 (1)~(6)の欄 に記載) ※注5	生物学的 特性の 方法 ※注5	データの 公開 期間 ※注6	共有・公開の有無 (データ処理済みのデータ) 無の場合、黒と記載	制限共有		制限公開 ※統計データの掲載は統計の対象となつた件数を記載					制限公開 ※統計データの掲載は統計の対象となつた件数を記載					制限公開 ※統計データの掲載は統計の対象となつた件数を記載											
											共有・公開の有無 (データ処理済みのデータ) 無の場合、黒と記載	制限共有 制限公開	制限公開 公開期間	H28年 度	H30年 度	H31年 度	H32年 度	H33年 度	H34年 度	H35年 度	H36年 度	H37年 度	H38年 度	H39年 度	H40年 度	H41年 度	H42年 度	H43年 度	H44年 度	H45年 度				
		研究開発から提出される データ及びデータ類のうち、 データシェアリングポリシー の対象となるデータの概要 ※注1 ※注2			解析データの 解析方法 (データ分類 (1)~(6)の欄 に記載) ※注3	解析データの 解析方法 (データ分類 (1)~(6)の欄 に記載) ※注4	関連する表現 で示した データ分類 (データ分類 (1)~(6)の欄 に記載) ※注5	生物学的 特性の 方法 ※注5	データの 公開 期間 ※注6	共有・公開の有無 (データ処理済みのデータ) 無の場合、黒と記載	制限共有 制限公開	制限公開 公開期間	H28年 度	H30年 度	H31年 度	H32年 度	H33年 度	H34年 度	H35年 度	H36年 度	H37年 度	H38年 度	H39年 度	H40年 度	H41年 度	H42年 度	H43年 度	H44年 度	H45年 度					
①		〇〇病感受性領域研究にお ける疾患群を対象としたゲノ ムデータ	血液		WGS	FASTQ	関連する表現 で示した データ分類 (データ分類 (1)~(6)の欄 に記載) ※注5	2,000人	〇年度	raw	processed	統計	AGD	〇病感受性領 域研究における 疾患群を対象と したゲノムデー タとその解析件 数	0	1,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000	2,000				
		〇〇病感受性領域研究にお ける疾患群を対象としたゲノ ムデータ					臨床情報(診断 疾患名、血液検 査、年齢、体 重)	1,200症例	〇年度	raw	processed	統計	AGD	〇病感受性領 域研究における 疾患群を対象と したゲノムデー タとその解析件 数	0	600	1,200	1,200	1,200	1,200	1,200	1,200	1,200	1,200	1,200	1,200	1,200	1,200	1,200	1,200				
②		〇〇病感受性領域研究にお ける臨床組織を対象と したゲノムデータ	臨床組織		GWAS	VCF		100人	〇年度	raw	processed	統計	AGD	〇病感受性領 域研究における 臨床組織を対象 としたゲノム データとその解 析件数	0	50	100	100	100	100	100	100	100	100	100	100	100	100	100	100				
		〇〇病感受性領域研究にお ける疾患群を対象と したゲノムデータ					臨床情報(診断 疾患名、血液検 査、年齢、体 重)	100症例	〇年度	raw	processed	統計	AGD	〇病感受性領 域研究における 疾患群を対象と したゲノムデー タとその解析件 数	0	50	100	100	100	100	100	100	100	100	100	100	100	100	100	100				
①														実質																				
②														実質																				
③														実質																				
④														実質																				
⑤														実質																				

(注1) 連結させるデータセットは同じ番号を記載
(注2) データの分類、データシェアリングポリシー II.1. に示される(1)~(8)のうちかきを記載
(注3) データの解析方法: HGS, GWAS, WES, SNP array等を記載
(注4) 解析データのフォーマット: FASTQ/BAM/VCF, その他(その他の場合のデータ名称)等を記載
(注5) 人・年齢・性別・種別のいずれかを記載