

厚生労働行政推進調査事業費補助金（厚生労働科学特別研究事業）  
分担研究報告書

機微なデータの公開と共有  
ANDS ガイドからの機微なデータの公表と共有に関する主要メッセージ

木村 映善 国立保健医療科学院保健医療情報管理分野 統括研究官

## 研究要旨

国立研究機関所属の研究者のアンケート結果から、機微な個人情報を含むデータは公開することが困難ではないかという意見が散見された。オープンデータと言え、データを一般に広く公開するという印象があるが、最終的な理念はデータの二次利用の機会を最大化することにある。すなわち、データをただちに公開できなくても、そのデータの存在を知らしめることに意義がある。ANDS ガイドラインは機微なデータを公開するにあたっての方法論を紹介している。本稿はそのガイドラインを抄訳し紹介するものである。

- 簡単に適切な手順を踏めば、機微データを公開することの利点は、潜在的な不利益を遙かに上回る可能性があります。
- データを公表する、あるいはデータの記述（つまりメタデータ）だけを公開することによって他の人がそのデータを発見して引用できるようになります。
- データ自体をオープンにアクセスできるように公開せずとも、データの説明のみを公開できます。
- 公開されたデータへのアクセスに条件を付けることができます。
- 非特定化されたセンシティブなデータを公開できます。

### A. イントロダクション

#### A.1 このガイドについて

“ジャーナルや助成機関の方針でデータの共有が義務付けられているにもかかわらず、データをどのように共有すべきかについての実践的なガイダンスはほとんどありません。

んでした...” (Hrynaszkiewicz et al. 2010[1])

本ガイドは、オーストラリアにおけるセンシティブな研究データの公表と共有に関するベスト・プラクティスの概要を示したものである。本ガイドは、「センシティブデータの公開と共有ポスター（第 1.2 節）」で概説されているように、センシティブデータの公開と共有に必要な一連の手順に従っている。このポスターに記載されている手順の詳細と文脈を提供している。興味のある方のために、さらなる読み物のための参考文献が提供されている。以下のセクションとその中のステップに従うことで、データを安全に共有することについて、明確で合法的、倫理的な決定を下すことができるようになります。ほとんどの場合において可能です。

ANDS ガイド<sup>1</sup>とセンシティブデータ共有の成功事例の紹介

Sarah Olesen 博士は、センシティブデータとその公開に至るステップを定義している。Leigh Tooth 准教授の女性の健康に関する

<sup>1</sup> 原著の名前は Australian National Data Service

Guide。以下 ANDS ガイドとして言及

る縦断的研究が、どのようにして 20 年間にわたって健康と社会のデータを安全に共有し、大成功を収めてきたかを説明している。

この話は YouTube で視聴できる。  
<https://www.youtube.com/watch?v=5j9ftReZc5g&feature=youtu.be> [38mins]。

このガイドとあなたの機関の方針との相互作用

このガイドは、データ管理と公開に関する機関の方針をサポートすることを目的としている。ほとんどの研究者は、所属機関の方針や資金調達の取り決めの中で活動しているため、データの公開に関する決定がこれらの方針に沿ったものであることを確認しなければならない。特に知的財産に関連しており、時にはセンシティブなデータの分類（例：NSW 政府環境・遺産局のセンシティブなデータ種のポリシー）やデータレポジトリの選択にも関連している。

## A.2 機微なデータ公開と共有についてのポスター

オリジナルのポスターはこちらで公開されている。

[http://www.andis.org.au/\\_data/assets/pdf\\_file/0010/385309/sensitive-decision-tree.pdf](http://www.andis.org.au/_data/assets/pdf_file/0010/385309/sensitive-decision-tree.pdf)

翻訳したポスターは別途報告書に同梱している。

「ANDS データ GL オリジナル版」は、上記のオリジナル版を翻訳したものがある。

「ANDS データ GL 国内制度対応版」は、上記のオリジナル版の翻訳をベースに、木村が国内制度との整合性を配慮して加筆修正した案である。下記、国内制度対応版のポスター中の文言についての解説を記す。

「他者によって収集されたデータを使用しているならば、そのデータについて引用してください。」

この文言は、他の研究者・事業者が作成したデータを二次利用して新たなデータが作成された場合、そのデータを元に作成したことを明記することを求めている。これまでのデータの引用については、データを作成しているプロジェクトやWebサイト等の粒度での言及にとどまり、データそのものの引用ではないものが散見されていた。その理由として、当該データを利用している研究者であれば、どこにどのようなデータがあるかの知識は常識的なものであり、また筆者に確認すればいつでも対応できるという状況から、厳密にデータ引用をしていくというモチベーションはなかったようである。ただ、今後は、データ作成者に対して正当な評価を与えられるように定量的な評価指標を収集していく過程において、データ引用の標準化と集計可能な状態で記述をしていくことが不可欠であることから、引用についても今後は可能な限り標準的な手法で行うことを要請したい。

## A.3 何故機微なデータを公開、共有するのか？

### Box1. 用語の定義

データの公開と共有 (Data publication vs sharing) : 公開 (publication) とは、データが一般に公開される (made public) ことである。これには、一般公開されたデータについての説明と、データ自体へのアクセス、または条件付きアクセスに関する情報が含まれる。データの共有は、データを他の人が利用できるようにする場合をさす。常にデータの公開を伴うわけではない (例: 同僚間でデータを共有しているが、一般には公開されておらず、外部から発見できない場合や利用できない場合)。

メタデータ (Metadata): データに関する

データ。メタデータの一般的な例としては、図書館のカatalogレコードがある。

リポジトリ (Repository) : メタデータやデータが保管されている場所。データリポジトリには通常、一般の人がデータを検索したり発見したりできるオンラインポータルがある。機関別、分野別、一般的なリポジトリがある。

非特定化データ (de-identified data) : データの人や対象者が特定されるリスクを取り除くために、修正されたデータ。

機微性の高いデータは、データの公開と共有に関する議論からしばしば除外されてきた。機微性の高いデータの共有は倫理的ではない、あるいは安全に行うことが難しくすぎると考えられていた。この見解は、データを「脱機微化' de-sensitise' 」(すなわち非特定化) する方法の理解と利用が進むにつれて変化してきた。出版社や資金提供者が研究者にデータを公表し、共有することを求めるようになったことで、機微性の高いデータの共有も増加している[2 3 4 5]。オーストラリアやその他の国には、機微性の高い情報を含む重要かつ質の高いデータセットが存在する。これには、健康に関する疫学調査、医療試験データ、絶滅危惧種の生態学的研究などが含まれる[6 7]。そして、ほとんどの場合、被験者を特定できるようなデータセットの一部を削除したり修正したり、データへのアクセスや再利用に条件を付けたりすることで、このようなことが可能となる。データを共有することの利点は、出版社や資金提供者の要件を満たすだけではない。研究者、研究機関、研究参加者にとってのメリットは明らかである。

・データやデータセットの説明が公開されれば、他の人が発見することができるので、引用したり、元のデータ収集者や所有者への帰属を表明させたりできる。

・公開されたデータが添付されている学術論文は、それが無い論文よりも引用される頻度が高いというエビデンスがある[8]。

・あなたのデータを他の人と共有して再利用することを選択した場合、新たな共同研究や出版物が生まれる可能性がある。

・データを利用できるようにすることで、高価なもの、稀少なもの、再現性のないものの再分析が可能になる。

・医学研究のデータを共有することで、不必要な重複を減らし、これらの研究から得られたデータを最大限に活用することで、臨床研究の被験者の負担を軽減することができる。

・現在の研究環境や変化する研究環境では利用できないような、安全で継続的な保存が可能でストレージを提供する公開リポジトリにデータを保存できる。

## B. センシティブ (機微) データとは？

### B.1 センシティブデータの定義

センシティブデータとは、個人、種、物、プロセス、場所を識別するために使用できるデータであり、差別、危害、好ましくない注目を引き出すリスクをもたらすものである。殆どの法律と機関の研究倫理ガバナンス下では、センシティブなデータは、例外を除いて、通常、この形式で共有することはできない。

### B.2 人に関するデータ

センシティブな人に関するデータとは、一般的には機微な個人情報をさす。個人情報は、個人またはグループを識別するために使用できるものである[10 11 12]。個人情報は直接個人を特定できる情報 (氏名、生年月日、出生、住所など) と、表1の1つ以上の情報が伴った場合、センシティブなものとなる。

表 1. 1988 年のプライバシー法 (Part II, Division I, Section 6) で定義されたセンシティブ情報の種類

<ul style="list-style-type: none"> <li>・人種、民族的出自</li> <li>・政治的意見</li> <li>・政治団体の会員</li> <li>・宗教的信念または所属</li> <li>・哲学的信念</li> <li>・専門家または業界団体の会員</li> <li>・労働組合の会員</li> <li>・性的指向または慣行</li> <li>・犯罪歴</li> <li>・健康情報（健康情報の定義については、セクション 6FA を参照）</li> <li>・遺伝情報</li> <li>・生体認証情報</li> </ul>
---

アボリジニとトレス海峡の人々のデータ共有と実践に関する専門的な情報は、以下に記載されている。

- ・ AIATSIS Collection Access and Use Policy
- ・ AIATSIS Guidelines for Ethical Research in Australian

### B.3 人間のデータを越えて

センシティブデータには、希少種、絶滅危惧種、商業的に価値のある種の所在を明らかにするデータや、その他の保全活動を含むこともある [13]。合意されたセンシティブな環境・生態学的データの定義の必要性についてはよく説明された資料がある [13][14]。オーストラリア政府の環境省は、センシティブの程度は、脅威の種類、脅威のレベル、脅威に対する脆弱性、情報の種類に照らして考慮する必要があると指摘している。

センシティブな生態学的データの定義が明確になっていないのは、この分野ではデータのセンシティブさが時として過渡的な性質を持つことがあるためである。例えば、

カエルの個体数が少なく、ある場所では減少しているにもかかわらず、別の場所では減少していない場合があるときは、前者ではセンシティブが高く、後者ではセンシティブが低いと考えられる。生物多様性グローバル情報ファシリティ (GBIF) の Chapman and Grafton (2008) [14] は、センシティブなデータを「一般に公開された場合、種や保全活動に「悪影響」を及ぼす」情報と定義している。特に、このデータの公表や共有が、危害のリスクを増大させたり、種に悪影響を及ぼしたりする結果となる可能性につながるような情報を意味する。この GBIF レポートはさらに、データ所有者が、データセットのどの要素がこの種のセンシティブに抵触するかを明確にし、その決定事項をデータセットのメタデータに明確に文書化することを推奨している。

### B.4 センシティブな生態学的データの記述

ニューサウスウェールズ州のある地域で、数種の鳥類の個体数と繁殖習性を推定するための調査が半年間にわたって実施された。この調査で得られたデータには、各種の地理的データと繁殖状況が含まれている。今回の調査対象種の一つであるグロッシー・ブラック・オウムは、ニューサウスウェールズ州環境遺産局の「センシティブな種リスト」において「脆弱な」種としてリストアップされている。このリストによると、オウムの脆弱性の理由は「採卵と巣の攪乱の危険性」であるとされている。また、種の「センシティブ」の原因を特定することで、データの所有者はデータセットのどの要素を削除したり修正したりして、データセットの識別を解除できるかを特定できる (Section 5)。

チェック： あなたのデータはセンシティブですか？

あなたのデータに表1の情報、秘密なまたは神聖な慣習 (sacred practice) が含まれている場合、または情報が公開された場

合に種に悪影響を及ぼす結果となるような情報が含まれている場合は、センシティブである可能性がある。データの内容そのものを検討することに加えて、もしデータが共有された場合、そのデータが差別の原因となる可能性があるれば、そのデータはセンシティブなものとなる。後者のこの観点は、データ公開に関する倫理的な考慮事項に関連している（セクション5を参照）。

## B.5 文脈に応じた機微性

データの中には、生来的に機微なものもあれば、機微になっていくものもあれば、「機微である」とみなされてしまうものもある。

何が機微であると考えられるかは、時間の経過とともに、また人や被験者のグループ間で異なるかもしれない。機微でないデータは、その文脈によって、あるいはより多くの情報が追加されたときに機微になることがある。組織や出版物の購読者リストの名前は、通常、機微性があるわけではない。しかし、そのグループの特別な関心が購読者を差別にさらす可能性がある場合には、このリストは機微なものとなる可能性がある[12]。例えば、極端あるいは人気のない政治的団体などである。生態学的データの場合、ある種について、ある地域では人間の活動によって被害を受ける危険性があるが、別の地域ではそうではない場合がある。後者の場合、機微性は種そのものではなく、データが参照している場所によって決定されるかもしれない。明らかに機微ではない（すなわち、名前や生年月日が含まれていない）データ、または非特定化されたデータが、文脈が変われば再び機微になる場合もある。2つの一般的な例がある。

1. Triangulation：参加者または被験者の身元または機微な情報が、機微性の低い情報をいくつか組み合わせて特定できる場合。

例えば、人に関するデータでは、年齢、職業、家族構成などの情報があれば、比較的小さなサンプルでも、その人を特定することは難しくない可能性がある。

## 2. データのリンケージ

データのリンケージとは、同一人物や研究対象を含む2つ以上のデータセットを組み合わせることである。データセット単体だけでは、個人を特定したり、被験者を危険にさらしたりするのに十分な情報を含んでいないかもしれないが、2つ以上のデータセットと組み合わせることで、個人を特定可能になるかもしれない。例えば、データセットAは、認知障害のある、特定できない患者のグループの病歴を記述している。データセットBは、同じく中程度に小さな人口での雇用情報と公共交通機関の利用状況を含む。これらのデータセットがリンクされると、患者がどこで働き、どこに住んでいるかについての十分な情報があり、患者が特定される可能性がある。生態学研究からの例をあげる。データセットAは、採卵に弱いハヤブサの種と、その個体数と場所を経時的に記述している。データセットBは、同じハヤブサの種の同じ期間の繁殖パターンを含む。これらのデータセットがリンクされた場合、営巣時のハヤブサの位置を決定するのに十分な情報があるかもしれない。データの所有者や管理者は、常にデータセットの中でTriangulationの可能性を検討し、その可能性があるかを確認する必要がある。新しいデータが導入されたときに、データのリンケージが行われる可能性をふまえてTriangulationを再考するのは良い習慣である。

C. 私は機微なデータを持っています - さて、どうしますか？

## C.1 機微なデータの公開と共有方法

前のセクションでは、元の形式の機微データについて説明した。ANDSの「非特定化のためのガイド」では、元のデータを修正

して機微性が残らないようにする方法を説明している。この種の修正されたデータは、研究参加者や被験者を危険にさらすことなく、他の研究者にも再利用可能であり、非常に価値のあるものである。また、そのデータは、より少ない法的・倫理的制約の下で公開・共有できる。ほとんどの場合、非特定化された機微なデータは条件付きアクセスで公開される。データに関する記述（すなわち「メタデータ」）がデータリポジトリで公開されるので、制限なくデータの存在が発見される状態になる。そして、データ自体は、ある条件（研究者やリポジトリによって設定された）が満たされるとアクセス可能になる。公開されたデータの説明には、アクセスに関する情報と、アクセスを申請する方法や場所へのリンクが含まれている。

## C.2 何が合法か？

オーストラリアのプライバシー法（Australian Privacy Act 1988[10]）の下では、機微性の高いヒトや個人のデータは一般的に元の形のままで共有できない。しかし、いったん非特定化された後は、これらの修正されたデータはもはや同法の対象とはならない。言い換えれば、非特定化された機微なデータは合法的に共有することができる。1988年オーストラリア・プライバシー法では、非特定化の定義を「個人情報、情報が識別可能な個人または合理的に識別可能な個人に関するものでなくなった場合に非特定化される」としている（第2部第1部第6節）。プライバシー法はまた、「識別情報」の定義を提供しており、個人をデータから識別特定できないようにするためには、最低限どのような情報を削除しなければならないかを示している。プライバシー法（1988年）は、識別情報が除去されたデータには適用されないが、データを非特定化する行為（すなわち、元の機微性の高いデータセットから識別情報を除去する行為）には適用されることに注目すべき

である。しかし、この行為は、プライバシー法（1988年）のオーストラリアのプライバシー原則の中で、機微性の高いデータの利用に対する数少ない例外の一つとして明示的に容認されている。これは、特定の同意を必要とせずに「個人が自分の個人情報が使用されたり開示されたりすることを合理的に期待することができる」「通常の...慣習」と考えられているからである[16]。データ共有における同意の詳細については、「倫理的配慮」のセクション（セクション5）を参照すること。

機微なデータの共有の合法性に関する重要なポイント

- 機微性の高いデータは、ほとんどすべての場合、元の形で公開したり共有したりすることはできない。
- 非特定化されたデータは、もはや機微なものではなく、共有することができる。
- 人や個人のデータを非特定化する行為は、オーストラリアプライバシー法（Australian Privacy Act 1988）の下で認められている[16]。

## D. データの非特定化

データ共有のための非特定化の目的は、研究の参加者や被験者（例えば、動物や植物の種）が識別され、危害や差別の危険にさらされることを防ぐことである。これには、元の（機微性の高い）データセットの情報を削除したり修正したりすることが含まれる。削除したり修正したりする必要がある情報は、データセットの内容や、データが機微であるとみなされた理由によって異なるだろう。

チェックポイント：私のデータを非特定化することはできますか？

データの中には、参加者や被験者への危

害のリスクを十分に最小化するために修正できないものや、修正した場合でもそのデータの利用価値が損なわれてしまうものがある。これは、データが収集された文脈を配慮することで身元の特定可能になることが多いからである。例えば、踏み込んだ個人面接からのデータ、および アイデンティティがデータの本質的な側面である文化的または歴史的なデータなどである。前者の場合、データは研究に参加していない（したがって同意していない）他の個人の アイデンティティも明らかにすることがある。大多数において、参加者からの明示的な同意およびヒト研究倫理委員会の承認なしには、非特定化できないデータを公表することはできない。ただ、メタデータが研究に参加した個人または被験者を特定するために使用できない場合には、必ずしもメタデータの公開を妨げるものではない。

#### D.1 生態データ

環境データや生物多様性データの手法は、ヒトデータに比べて確立されていない。ヒトデータの場合、機微性は通常、個人を特定することに関係している。したがって、非特定化するためには、識別情報を除去しなければならない。しかし、生態学的データの機微性は、一般的には種の名前そのものではなく、その場所や繁殖・結実・移動の時期に関する情報に関係している。例えば、第2.4節の例示的なデータセットは、単にグロッキー・ブラックコ・ガトウのデータが含まれているからといって機微性が高いわけではなく、繁殖時の正確な位置に関するデータが付随しているからである。機微性が低下した生態学的データの実例は、オーストラリア政府環境省の「センシティブな生態学的データ - アクセスと管理に関する方針 V1.0」[15] の表1にある。

#### D.2 再同定のリスク管理

データが他のデータとリンクされている場合、データが非特定化された後に、再特定のリスクを見直さなければならない。データリンケージとは、同じ人や被験者のデータを含む 2 つ以上の別々のデータセットをマージすることである。疫学、医学、社会科学、生態学の分野ではますます一般的になってきているが、その理由は、研究者が新たなデータを収集しなくても、より多くの情報を追加することで、人々や被験者の文脈をより詳細に理解できるからである。また、データリンケージは、既存のデータセットからより大きな価値を引き出すことができる。Triangulation と同様に、データリンケージは、（新たにリンクされたデータによって）1 つまたは複数の潜在的に識別可能な情報が追加されたために、識別されていなかった参加者や被験者が再特定される可能性があることを意味している。データがリンクされる際には、新しいリンクされたデータセットを（潜在的に）識別可能なデータセットとして扱い、そのリスクを評価することで、この再特定可能性を評価しなければならない。

#### D.3 いまのうちに

公開のためにリポジトリに提出するデータセットは、以下のようにしなければならない[1]。

- ・クリーンアップ; エラー、外れ値、重複、欠落データをチェックする（本当に欠落しているデータにはそのような注釈をつける）
- ・注釈をつける; 変数とオブジェクト（関連する場合はそれらのカテゴリ）には明確なラベルをつけて記述し、必要に応じてキーを提供すべきである。変数やオブジェクトが非特定化の間に元の形から変更された場合は、そのことをメモしておく。
- ・仕様がオープンであり、変換やアーカイブが容易なフォーマットであること。一般的に使用されている様々なオペレーティ

ングシステムやプログラムで読みやすく、操作しやすいデータ形式を使用するように努める。アクセシビリティを高めるために、非専有（「オープン」）フォーマットも推奨する。データ形式に関する情報は、メタデータレコードの中で提供することが望ましい。データを読み込んで分析するために特殊なプログラムが必要な場合は、可能であればデータと一緒に提供すべきである。データフォーマットに関する特定の要件については、選択したデータリポジトリを確認する。

## E. 倫理的配慮

法律や地域社会の基準を満たすことに加えて、研究者は参加者や研究対象者に対して倫理的な義務を負っている。これには、プライバシーを保護し、研究への参加とその後の発表から生じる可能性のある危害を回避することが含まれる。参加者の信頼と研究の完全性を維持するためには、データの倫理的な管理が研究者の第一の関心事でなければならない。

### E.1 新規ヒトデータの場合、または研究参加者との接触が可能な場合

データを取得して公開する前に、ヒトの参加者の同意が必要である。同意を得るための最良のタイミングは、データが収集される前である。これは最良の実践であるだけでなく、研究プロセスの後半で同意を得ようとすることで、費用、遅延、データ利用の損失を回避できる。データの共有が要求された場合、参加者が研究への参加を拒否するだろうという懸念は、根拠のないものである可能性が高い。[17 18]

研究参加者に非特定化データの公開と共有についての同意を求めるにあたり、含めべき事項は、

1. 研究への参加に同意する前、およびデータ収集への同意を求められる前に、参加者に提供される「情報シート」に、データの非特定化、公開、共有のプロセスに関する情報を含める。この情報は、参加者にとって理解しやすく、提供したデータの公開および／または共有に同意することについて十分な情報に基づいた決定を下すことができるように、十分な詳細が記載されていなければならない。参加者情報シートは、研究のあらゆる側面が開始される前に、ヒト研究倫理委員会（HREC）の承認を得なければならない。それには、機微保持の手順、データの公表、データ共有が可能な条件（研究を実施している者以外の研究者とデータを共有する見込みがあるかどうかを含む）を（簡潔に）記載する必要がある[6]。

2. データの共有および／または公表に対する同意を「同意書」内で具体的に要求する。同意書においてデータの公開と共有を依頼する際の文言例を以下に示す。

3. データの公表と共有に関して、管轄区域内の倫理申請書、情報シート、同意書の要件を熟知しておく。例えば、National Health and Medical Research Councilのヒト研究倫理申請書（HREA）では、データの共有に関連したいくつかの質問、特に Q3.13～Q3.17が挙げられている。

### E.2 データの公開・共有を求める同意書

参加者の同意書におけるデータの公開および共有に関する記述は、以下の通りである。

1. データの非特定化、公開、共有を妨げることを避ける[19]。
2. 将来のデータ公開（リポジトリへの保存を含む）と共有の可能性を述べる。
3. データへのアクセスが他者に認められる条件を記述する。これには、匿名化するプロセスや、元の研究チームの承認などの他



の条件が含まれる可能性がある（セクション G.1「条件付きアクセス」を参照）。場合によっては、参加者が自分のデータを誰と共有することに同意するか（誰と共有しないことに同意するか）を選択する機会を提供することも適切な場合がある。

4. 収集したデータを文書化し、後のデータ利用者が、参加者が同意した条件を認識できるようにする。

### E.3 データの公開と共有を求める同意書の例文

データが公開されることを意図している場合、または少しの制限のみでアクセスできる場合：’

この研究の情報は、あなたが誰であることを明らかにしない方法をとおしてのみ使用されます。この研究の出版物や他の研究者と共有されるデータファイルでは、あなたの身元が特定されることはありません。貴方がこの研究に参加していることは秘密にされます。[20]

ファイルが他の研究者と共有されたり、結果が公開されたりする前に、あなたを特定できる個人情報は削除されたり変更されたりします。[18]

私は、研究のために収集された研究データが、私の名前やその他の識別情報が使用されない限り、公表される可能性があることに同意します。

条件付きアクセスが可能なデータの場合：

他の真正な研究者は、このフォームで要求された情報の機微性を保持することに同意した場合に限り、このデータにアクセスすることができます。[21]

上記の例は、その再利用時に適用される特

定のアクセス条件を含むように適応させることができる。

他の真正な研究者が将来的に非特定化されたデータへのアクセスを要求する可能性があります。このフォームで要求された情報の機微性を保持することに同意した場合にのみアクセスが許可されます。また、そのような研究者のアクセスには、元の研究チームの承認も必要となります。

また、参加者に、データ共有に同意（あるいは反対）する人を選択する機会を与えることも検討してください。例えば、データを再利用する可能性の高い人のリストから選ぶことができます。

### E.4 既存データの場合、研究参加者との再接触ができない場合

センシティブなデータは、以下の場合には、研究参加者の明示的な同意なしに合法的に共有することができる。1. データ収集の同意に先立って参加者に提供された情報が、データの将来の使用を示唆している場合\*、または  
2. 同意を得る機会がもはや存在しないか、または現実的ではない場合において、データは非特定化が可能であり、かつ非特定化のプロセスがプライバシー法（1988年）に規定された定義に合致している場合、かつデータの公表または共有が研究参加者または被験者に害を与えたり差別を助長したりする危険性がない場合、最初のデータ収集時の情報シートおよび同意書が共有を排除していない場合。

注：\* 参加者の同意書がデータの公開や共有について具体的に言及しておらず（いずれも除外していないこと）、情報シートが言及している場合、プロジェクトへの参加自体に同意があれば、共有が可能である。これは、参加への同意が情報シートへの理解

と同意を意味するからである。

推奨: これらの要件を満たしているかわからない場合は、研究プロジェクト以外の同僚や最寄りのヒト研究倫理委員会に意見を求めることを勧める。

#### E.5 自分が収集していない機微なデータの共有

あなたがデータの所有者ではない場合、データの公開や共有についてはどうすればいいのでしょうか?つまり、あなたはデータの再利用者または「二次データ利用者」である。データの再利用は、疫学のような非特定化データを扱う研究分野では、すでに比較的広まっている。上記の質問は、学術論文と一緒にデータを公開することを求める雑誌の要求に応えるために、ますます一般的になってきている。ほとんどの場合、自分が収集していないデータは、そのデータの著作権を所有していないため、公開できない。例外は、データ所有者が再配布を許可するためにライセンスを与えたデータの場合である。

研究倫理に関する注意事項:元の研究者と同様に、二次データの利用者にも「データが責任を持って敬意を持って使用され、参加者のプライバシーが守られていることを保証する義務」がある[6]。したがって、データの再利用者は、元の研究者が参加者に指定したデータ利用条件、および再利用のためのアクセスが許可された際にデータ所有者または管理者が提示した再利用条件を守らなければならない。

#### F. データの引用

もしあなたがデータの再利用者であるならば、そのデータに基づいたすべての論文、プレゼンテーション、助成金申請の際に、そのデータの元の出典を参照したり、引用

したりしなければならない。これにより、データの所有者はデータの利用状況を追跡できるだけでなく、あなたの研究をこのデータセットに基づいた他の論文とリンクさせることができる。この方法の詳細については、ANDS Guide to Data Citationを参照すること。

推奨:

- 再利用するデータがライセンスされている場合は、データ共有に関するライセンスの条件に従うこと。(一般的には「再配布(‘redistribution’)」と呼ばれている)再配布が許可されている場合は、データを共有し、データ所有者に帰属させることができる。あるいは、他の人がアクセスを要求できるように、公開された研究の説明にデータの出所を引用することもできる。
- 再利用するデータがライセンスされていない場合は、データの所有者または管理者に連絡して、公開と共有に関する指示を求めること。

データを再利用する場合には、すべての学術成果物に元のデータの出典を引用する必要がある。

#### G. データを発見可能にする

リポジトリ内の公開メタデータ(public metadata)は、あなたのデータを他の人が発見できるようにするための最良かつ最も簡単な方法である。以下は、公開カタログでのデータに関する説明である。

「データ自体を自由に利用できるようにしないとしても、データの説明を公開することはできます。」

リポジトリからは公開するにあたり、どのようなメタデータを用意する必要があるかを説明される。ほとんどの場合、助成金

申請書やプロジェクト報告書、論文などで既に持っている情報である。公開されたメタデータは、他の人があなたのデータを見つけて引用することを可能にする。データ自体がメタデータと一緒に公開されていない場合（ただ、多くはメタデータ内のリンクを介して公開されている。以下の例を参照）、メタデータはデータへのアクセス方法に関する情報を再利用者に提供する。

機微性の高いデータを持つ研究者は、自分のデータを公開して共有したいと考えているが、公開やアクセスについてはまだ懸念を持っているのが一般的である。これは、条件付きアクセスでデータを公開することによって、大部分が対応可能になる。

#### G.1 データへの条件付きアクセス。条件付きアクセスとは何ですか？

条件付きアクセス(conditional access)は、メタデータが公開されている（すなわち公開リポジトリでメタデータが公開されている）が、データ自体へのアクセスは、あらかじめ定められた条件が満たされた場合にのみ発生する。これらの条件は研究者やデータ所有者、および/またはデータリポジトリによって設定される。これらの条件には、データの再利用者となる可能性のある人に以下のことを要求することが含まれる。

- ・ 登録および/または連絡先の詳細を提供すること
- ・ データをどのように使用、保存、または管理するかについての情報を提供する。
- ・ データセキュリティ、プライバシーの条件に同意すること。
- ・ 共同研究やその他の目的でデータ所有者から連絡を受けることに同意すること
- ・ アクセス料金を支払うこと
- ・ 元の研究参加者（ヒト）が同意した同意フォームおよび情報シートに記載されているその他の条件を満たしていること。

機微なデータを発見可能な状態にしたいが、条件付きでアクセスできるようにしたいと考える理由には、以下のようなものがある。

- ・ 再利用者が真正な研究者であることを確認するため
- ・ 再利用者がデータの機微性または安全な保管を維持することを認識し、同意していることを確認するため
- ・ 参加者同意書または研究倫理申請書で条件付きアクセスを指定した場合
- ・ 誰が何のためにデータを使用しているのか、あるいは何のためにデータを使用しているのか、ある程度の監視を維持したいとき

その他の条件付きアクセスの理由としては、以下のようなものがある。

- ・ データ（またはデータの一部）が禁輸下にある場合
- ・ コラボレーションができるように、誰がデータを利用しているのかを知りたいとき

他の人は、データを新たな価値ある目的のために利用できる。研究データは、データ作成者や管理者が（他者による利用に）価値があると考えるかに関わらず、広く普及させるのが理想的である。データの用途が将来的にどのようなものが存在するかを知ることは不可能である。

#### G.2 データへの条件付きアクセスが可能な公開メタデータの例。オーストラリア女性健康縦断研究(ALSWH)

「女性の健康政策と実践の発展を支援するためのエビデンスベースを強化するための国の研究資源としての ALSWH にとって、デ

ータの共有は基本的なものです。私たちは、私たちのデータを利用できるようにし、幅広い分野の研究者間のコラボレーションを奨励することに全力を尽くしています。学際的なグループ間でのデータ共有は、新鮮な視点と女性の健康に関する新たな洞察と知識を得るための機会を提供します。」

教授 Gita Mishra, ALSWH ディレクター

#### ALSWH のメタデータレコード

オーストラリア・データ・アーカイブ（データ・リポジトリ）に保管されている女性の健康に関するオーストラリア縦断的研究のデータの「利用条件」は、メタデータ・レコードで明確に説明されている。

データの利用について正式な要請があった場合には、共同研究者がデータを利用できるようにすることができる。データ利用の許可を得る必要がある。

### G.3 ライセンス

再利用を目的としたオーストラリアのデータはすべてライセンスが必要である。これには、非特定化されたデータセットも含まれる。ライセンスとは、データの使用方法を明確に定めた文書であり、元のデータ所有者に帰属する。ライセンスがないと、データをどのように再利用できるかが不明確になり、再利用する人の意欲をそぐことになりかねない。

ライセンスには様々な形態があり、再利用に関する制限が少ないものから多くの制限があるものまで様々である。データリポジトリの中には、独自のライセンス文書を持っているところもある（例：Australian Data Archive の「Access Categories」）。また、オープンアクセス、つまり無制限のアクセスを要求するものもある（例：Dryad）。

クリエイティブ・コモンズでは、再利用の条件が少ないものから多いものまで、さ

まざまなライセンスを提供している。クリエイティブ・コモンズのウェブサイトでは、簡単な手順でライセンスを適用することができる。

注意すべき点は、あなたのデータにオープンライセンスを適用したとしても、機微性の高いデータを公開したり、データの非特定化に代わるものとして機能したりするわけではないということである。機微なデータは、ライセンスがあっても機微なままであるため、参加者の同意なしに公表することはできない。したがって、本ガイドでは、データの非特定化がなされた後、または特定可能なデータを共有することに同意されており条件付きアクセスの取り決めによって利用可能になっている場合のみに、ライセンスを適用し、公表することを推奨する。詳細については、データ再利用のためのライセンスおよび ANDS Research のデータ著作権およびライセンスに関する FAQ を参照すること。

### H. 公開する権利

ほとんどの場合、データを公開する個人または機関は、データを公開するための適切な権利（著作権など）を保有していなければならない。ライセンスは、そのデータの権利者のみがデータに適用できる。

#### H.1 データの著作権が自分にあるかを知るには？

著作権の所有権と著作権放棄の例は、オーストラリアの教育機関によって異なる。著作権は知的財産の一側面であるため、所属機関または雇用主の知的財産ポリシーを調べることを勧める。それでも不明な点がある場合は、所属する研究室または研究サービス部門にアドバイスを求めること。より詳細な情報については、ANDS Guide to Copyright, Data and Licensing を参照すること。

## H.2 自分のデータはどこで公開すればよいですか？

多くのデータリポジトリから選択することができる。リポジトリの中には、メタデータのカタログのみを提供し、データの保存場所にリンクしたり、参照したりするものもある。また、メタデータをカタログ化し、データ自体を保存しているリポジトリもある（例：Australian Data Archive、Figshare）。リポジトリには、機関固有のもの、分野やコンテンツ固有のもの、一般的なものなどがある。国際的なリポジトリのリストはre3data.orgで見ることができる。

機微データを公開する場所を選ぶ際には、以下の点を考慮すること。

- ・ リポジトリがデータへの条件付きアクセスをどのように管理しているか、また、条件付きアクセスはリポジトリが管理しているのかデータ所有者が管理しているのか。
- ・ 機関、雇用主、資金提供者、出版社が特定のリポジトリの利用を義務付けているか、推奨しているか。

データをどこで公開するかについてのその他の考慮事項としては、以下のようなものがある。

- ・ あなたの研究分野では、データを公表する場所に関する慣習があるか
  - ・ メタデータとデータを同じ場所で公開するかどうか
  - ・ リポジトリが必要とするメタデータ
  - ・ リポジトリがデータを必要とするフォーマット
  - ・ 発生した経済的コスト
- ・ リポジトリがデータに一意の識別子を割り当てることでデータ引用の追跡を可能にするか（DOI 識別子等）

## I. 謝辞

著者は、本ガイドの執筆にあたり、グレッグ・ラフリン博士（ANDS 主任政策顧問）、バーデン・アプリアード博士（AusGOAL 国家プログラム・ディレクター）、マイケル・マーティン教授（オーストラリア国立大学財務・アクチュアリー研究・応用統計学 研究科、人文社会科学委任倫理研究委員会委員長、オーストラリア国立大学科学・医学委任倫理研究委員会委員長）、ジェフ・トランター博士（Environmental Resources Information Network、ERIN）に貴重な助言とコメントをいただいたことに謝意と感謝を表したいと思います。

## J. Reference

原著を参照のこと

この文書は、Creative Commons Attribution 3.0 Australia License にもとづいて Australian National Data Service が公開した” Publishing and sharing sensitive data” を抄訳したものです。元ライセンスにもとづき、Creative Commons Attribution 4.0 International License にて公開いたします。

オリジナルの文書は、こちらからアクセスできます。

[https://www.ands.org.au/\\_data/assets/pdf\\_file/0010/489187/Sensitive-Data-Guide-2018.pdf](https://www.ands.org.au/_data/assets/pdf_file/0010/489187/Sensitive-Data-Guide-2018.pdf)



このガイドラインを抄訳した理由は、「機微なデータであるから、オープンデータに供することはできない」という誤解を解くために、このガイドラインが有用であると判断したためです。直接公開できないものに

については、データの内容について説明するメタデータのみを公開し、データの発見性、引用、コラボレーションの可能性を向上させることが期待されます。

なお、元文書の趣旨を把握することに要点を置いているため、レイアウト等の厳密な再現は意図しておらず、図面の一部を掲載していません。引用等の際は原著を必ずご確認ください。