

ARIM事業におけるデータ登録と共有形態について

東北大金研分析コアARIM事業班

1. 基本的な考え方

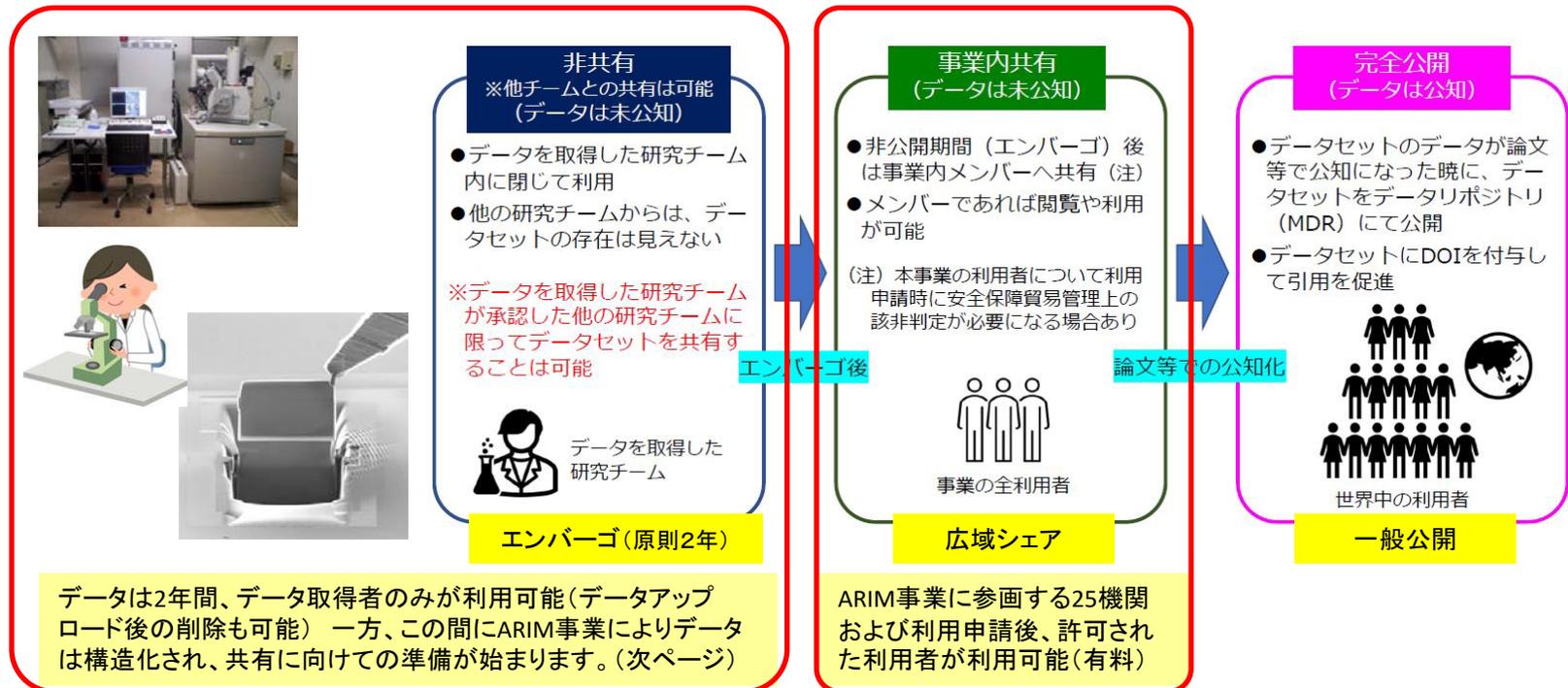
文科省ARIM事業では各種実験機器を国内外の研究者の方々に開放しているだけではなく、それらの機器を利用することによって得られたデータを一定期間の後に共有させていただくことによって、材料の研究開発を推進していくことを事業の一つの目的としています。

もちろん貴重な実験によって得られたデータは研究の根幹をなすものですから、データ共有に至る過程には一定のルールがあります。たとえば実験後、直ちにそのデータが公開されるわけではありません。また、本事業で定義するデータは登録データと事業側がデータ駆動型研究を推進するために構造化処理したものに大きく分けられ、それらの共有形態には時系列的な階層もあります。

データの利用権限や利用形態に関しては、データ登録者と事業側に約束が必要です。そのための基礎となるのがデータ登録約款*です。この約款の基本的な内容は事業に参画している25機関間で統一されている必要があり、東北大が利用者の方と交わす約款もARIM事業全体でまとめられた骨子を本学産学連携部の確認の上、ナノテク融合技術支援センター企画運営委員会で承認されたものです。ここではまず、事業全体におけるデータ登録や利用権限の概要をまとめ、後半では東北大計測分野におけるガイドラインを説明します。

(*「国立大学法人東北大学 マテリアル先端リサーチインフラデータ登録約款」(令和5年3月30日、東北大学ナノテク融合技術支援センター制定)

2. データ共有の形態と利用権限の大きな流れ



データ登録に同意していただいた場合でも、事業内で共有されるまで2年を要し、一般に公開されるのは、登録者がその旨を申請した場合です。

3. データの種類と権利および利用範囲の概要 (約款 第1条、9条、15条)

一口にデータといっても生データもありますし、ARIM事業側が広く利用者に提供することを念頭にフォーマットを変えたデータなどもあります。

3.1 登録データ: 装置利用者が登録するデータ (技術代行の場合は事業参加者が登録行為を行います、「データ登録者」は本来の利用者です。)

- (a) 生データ: 共用装置から直接得られるデータです。
- (b) 提供データ: データ提供される方が所有しているデータをさします。
- (c) 入力データ: 上記のデータに関する付帯情報などの、いわゆるメタデータです。

生データ(および提供データ)に関する著作権その他の権利は、当該データの権利者に帰属し、データを登録することにより第三者に譲渡されるものではありません。(約款 第9条1)

登録データに対しては、データ利用者は広域シェア以降であっても共有することができません。本事業従事者は広域シェアとなった段階で、事業の運営やサービス向上の目的に限り、アクセス等が可能となります。(第15条一)

3.2 構造化データ: ARIM事業側が第三者が利用しやすい形式に変換したデータ

- (a) 機械可読化データ: 生データを本事業機関が作成もしくは購入したプログラムによって変換したデータ
- (b) グラフおよび表データ: 機械可読化データに変換されたデータからグラフや表に変換したデータ
- (c) 選定メタデータ: 登録された方が入力した測定機器や材料に関するデータに対して用語の統一などを行ったうえで抽出したデータ
- (d) データセット: 上記のデータ群を一体化したもの
- (e) データカタログ: データセットの概要が抄録としてまとめられたもの

上記の構造化されたデータに関する権利は、本学に帰属します。構造化データに対しては、データ登録者の独占的利用権限は保証されないことにご注意ください。(約款 第9条2)

非共有期間中はデータ登録者のみが構造化データにアクセス可能です。またこの期間は登録データを含めて、削除可能です。広域シェアになった段階で、本事業機関は構造化データにアクセスでき、また第三者への提供が可能となります。(第15条)

4. データの共有形態 (約款 第14条、第15条十五)

- (a) 非共有期間: 原則2年間は登録および構造化データに対するアクセス権はデータ登録者のみが有します。(延長可能、原則1年、国プロ等は任意)
- (b) 広域シェア: 構造化データに対するアクセス権が事業側と許可されたデータ利用者に生じます(後者は有償)
- (c) データ中核拠点共用: データ登録者の申請により、構造化データがデータ中核拠点に複製・移転することができます。

5. データを登録する際に ARIM事業側にお認め頂くこと (許諾事項、(約款 第12条))

データをご登録いただくことにより、次の利用をお認めいただくこととなります。(約款 第12条二)

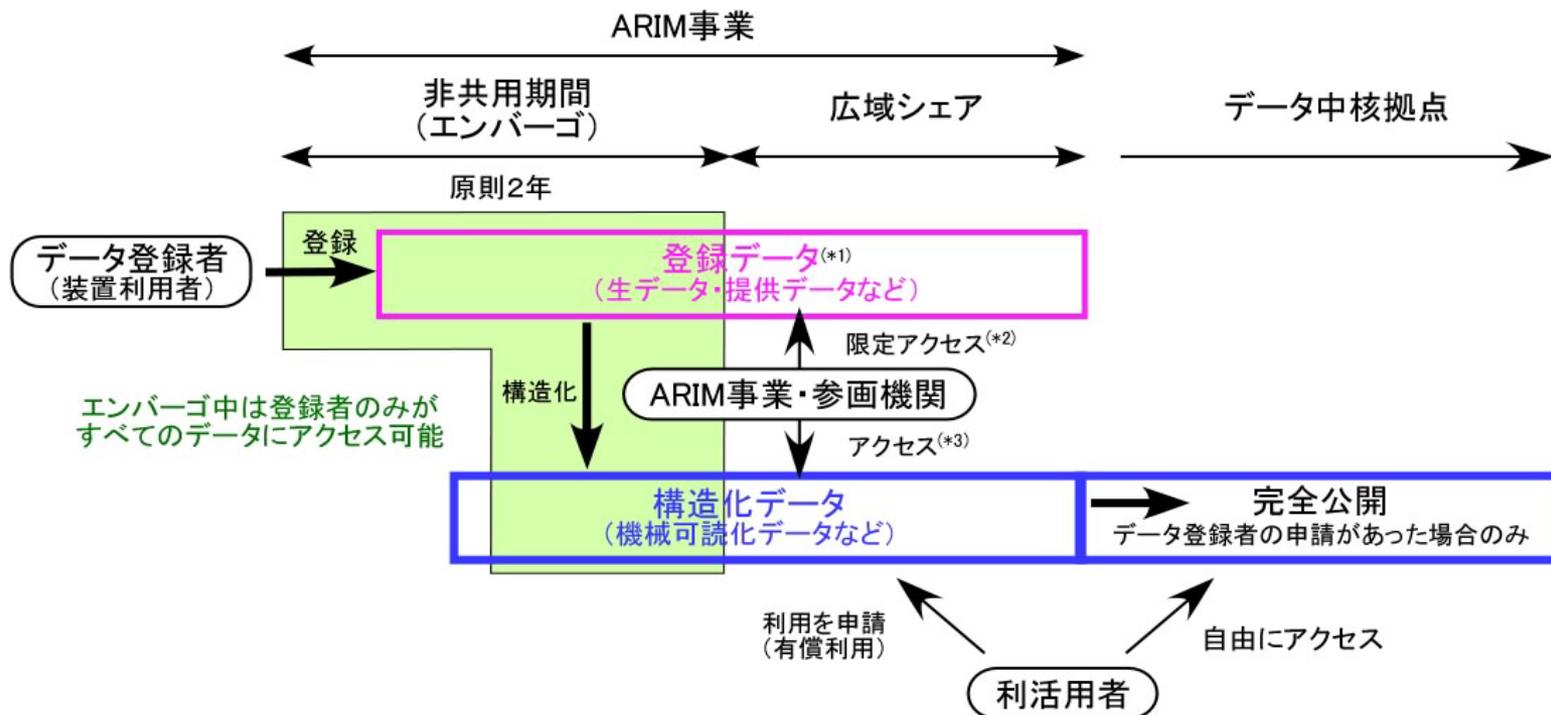
- (a) 登録データを構造化データとすること。
- (b) 登録データ・構造化データを本学もしくは本事業機関のデータベースに収納すること。
- (c) 広域シェア後の事業者ならびに事業が認めた利用者に対する構造化データを利用する権利。

6. その他の事項

- (a) データ登録者は次の事項を非表示とできます: (i) 課題番号、(ii) 氏名、(iii) 所属機関 (約款 第17条)
- (b) ARIM事業終了後、上記 3.1, 3.2 記載のデータを、本事業機関は新たな機関等に継承できます。(約款 第26条)

7. データの種類と利用権限のまとめ

ARIM事業で统一的に整備され、各実施機関において施行された約款の内容のうち*、データ利用権限に関することは下図のようにまとめられます。



*1: 登録データはデータ利用者に共用されません。

*2: 事業の運営やサービスの向上等の事業に資する目的に限り、事業側は登録データを閲覧、検索、編集、ダウンロード、二次利用することができます。

*3: 広域シェアの構造化データに対して、事業側は閲覧、検索、編集、ダウンロード、二次利用および第三者への有償・無償による提供ができます。

* 装置ご利用者のみなさまには課題申請時にデータ登録約款の内容をご承諾していただくことになります。

8. その他

8.1 構造化データの形態について

⇒ ARIM事業で展開しているデータ構造化は現時点(2023.5)ではフォーマットの統一の段階にあり、機械学習によるデータ駆動型研究への展開はこれから行われる予定です。したがって、構造化されたデータの形態は当該事業としてこれから精査・改良されていくことをご了解ください。

8.2 利活用制度

⇒ データ利活用の制度設計は最も早く非共用期間が始まる2年後を目指して、検討が開始された段階で、データ利用料金やデータ利用に関する取り決めなどは決定次第、アナウンスいたします。