

# データベースレスキューマニュアル

Ver. 1.0  
2021 年 3 月

JPCOAR 研究データ作業部会  
データベースレスキュープロジェクト

天野絵里子：京都大学学術研究支援室  
岩井雅史：信州大学附属図書館  
大園隼彦：岡山大学附属図書館  
木下直：鳥取大学研究推進部図書館情報課  
末田真樹子：国立情報学研究所学術基盤推進部学術コンテンツ課  
中竹聖也：東京大学情報システム部  
中筋知恵：小樽商科大学学術情報課  
南山泰之：国立情報学研究所オープンサイエンス基盤研究センター  
結城憲司：北海道大学附属図書館研究支援課



この文書は [クリエイティブ・コモンズ 表示 4.0 国際 ライセンス](https://creativecommons.org/licenses/by/4.0/) の下に提供されています。

# 目次

1. はじめに.....	5
1-1. データベースレスキューとは.....	5
1-2. 本マニュアルについて.....	5
1-3. データベースのライフサイクル.....	6
2. データベースのキュレーション.....	7
3. レスキューの前準備.....	7
3-1. 研究者・関連部署等との交渉.....	8
3-2. 対象となるデータベースの現状調査と受入先の検討.....	8
3-2-1. システム、ファイル形式、データ数・ファイル容量、ハードウェアの確認.....	8
3-2-2. 受入先のサーバ・システム.....	9
3-3. レスキュー計画の策定.....	9
4. レスキューの手順.....	9
4-1. データの取得.....	10
4-1-1. 取得方法を決定する.....	10
4-1-2. ウイルススキャンの実施.....	10
4-2. データの確認.....	10
4-2-1. データの内容を確認する.....	10
4-2-2. データ容量、データファイル形式を確認する.....	10
4-2-3. データベースに関するドキュメントを確認する.....	11
4-2-4. データ保存の粒度を決定する.....	11
4-2-5. 個人情報、機密情報等の公開できない情報の有無を確認する。.....	11
4-3. メタデータの検討.....	12
4-3-1. メタデータを抽出する.....	12
4-3-2. メタデータスキーマを決定する.....	12
4-3-3. メタデータをマッピングする.....	12
4-3-4. 必要に応じてシステムに新たなメタデータフィールドを追加する.....	12
4-3-5. DOI の取得.....	13
4-4. システムへの登録.....	13
4-5. 公開範囲、公開日、公開方法を決定する.....	13
4-5-1. 公開範囲を決定する（制限を設ける場合）.....	13
4-5-2. 公開するデータのライセンスを決定する.....	13
4-5-3. 公開方法.....	14
4-5-4. 公開日の決定.....	14
4-6. オリジナルデータベースの停止.....	14

5. レスキューしたデータベースの維持管理.....	14
5-1. 継続的なメンテナンス .....	14
5-1-1. 公開データのバックアップ .....	14
5-1-2. データ形式の変換作業（マイグレーション） .....	14
5-2. ドキュメントの整備.....	15
5-3. 利用統計のフィードバック .....	15
5-4. データの追加・修正.....	15
5-4-1. データ本体の追加 .....	15
5-4-2. メタデータの追加・修正 .....	15
5-5. 連絡体制の構築.....	15
6. データベースの取下げ・消去 .....	16
6-1. 取下げ・消去の検討が必要になるケース .....	16
6-2. 取下げができないケース.....	16
6-3. 取下げの手順.....	16
6-4. 取下げた後の処理 .....	17
6-4-1. 再公開の可能性を検討 .....	17
6-4-2. バックアップを取ったデータの扱い .....	17
7. データベースの持続可能性向上のための研究者向けアドボカシー .....	17
8. 今後の課題.....	18
8-1. メンテナンスコストと持続可能性.....	18
8-2. ライセンスの考え方.....	18
参考文献.....	19
付録：システムを含めてデータベースをレスキューする場合の技術的な課題.....	20
1. はじめに.....	20
2. データの取り出し .....	20
2-1. 現状の公開システムおよびデータファイル形式の把握 .....	20
2-2. レスキューすべきデータを確認する .....	20
2-3. データの抽出 .....	21
3. レスキュー先のハード面 .....	21
3-1. 既存のシステムにレスキューする場合.....	22
3-2 新規にシステムを構築する場合 .....	22
4. レスキュー先のソフト面 .....	22
4-1. どのようにシステムを構築するか.....	22
4-2. 構築にあたって留意すること .....	24
5. バックアップ体制.....	24
5-1. データの責任の所在を明らかにする .....	24

5-2. バックアップ方法 .....	25
5-2-1. バックアップする対象 .....	25
5-2-2. バックアップ手段 .....	25
5-2-3. バックアップの体制 .....	25

# 1. はじめに

## 1-1. データベースレスキューとは

オープンサイエンスが世界的に推進される以前から、学術的に価値のある多くのデータベースが研究者によって構築され、ウェブ上に公開されてきた。しかしながら、公開から時間を経たデータベースには、さまざまな理由によってその維持管理に問題を抱えるものが少なくない。中には維持が困難となって公開を停止したり、データベースシステムやデータの更新が滞ったりと、利用に深刻な支障をきたしているものがある。公的な資金を財源として創出された研究成果はなるべく多くの人々に公開して共有し、学術の発展を促進しようというオープンサイエンスの考えに即せば、存続の危機にあるデータベースも、適切にその価値を継承していくことが求められる。

以上の背景を踏まえ、研究者による維持管理が難しくなったデータベースを機関リポジトリやデジタルアーカイブなど研究機関が管理する基盤やウェブサーバの上に適切に「レスキュー」し、将来的な利用環境を提供することもオープンサイエンスを推進する一つの取り組みとなりうる。そのような取り組みを「データベースレスキュー」と呼んでいる。また、広義では、アナログデータのデジタル化というプロセスを含めてデータベースレスキューと呼ぶ場合もあるが、本マニュアルではもともとデジタルであるデータベースのみを扱っている。

なお、データベースの検索機能等も含めてレスキューするのではなく、データベースに収められたデータのみのレスキューを行う場合もあるが、本マニュアルでは後者も「データベースレスキュー」の活動に含めている。

## 1-2. 本マニュアルについて

本マニュアルは、なるべく多くの研究機関の機関リポジトリやデジタルアーカイブの担当者に活用されることを期待し、昨年度までに本作業部会で収集したレスキュー事例を踏まえ、データベースレスキューを適切に実施するための手順や考え方を一般化して示している。また、解決方法が複数ある場合は選択肢を提示するかたちでまとめている。加えて、データベースの維持管理が困難になる前に、どのような情報を研究者に対して提供すればよいか担当者が検討し、情報提供できるようになることも想定して構成している。

ただしこのマニュアルは、必ずしも困難な状況にあるデータベースをレスキューする場合のみを想定しているわけではない。上記にあるような、維持管理に困難が生じたデータベースを移行する事例が実際には多いため、わかりやすくレスキューの場面を想定しているが、研究者が作成して維持管理していた学術的価値のあるデータベースを、その維持管理状況によらず、研究機関の基盤的機能（機関リポジトリやデジタルアーカイブ）に移行して継承する場合も含めた内容となっている。

また、付録として、機関リポジトリやデジタルアーカイブなどの既存の基盤への移行ではなく、データベースをシステムごとレスキューする場合についての課題をまとめている。付録としたのは、実際にシステムごとレスキューに取り組むためには受け入れ側に相応のリソースが必要であることや、事例によって対応が異なるため、現時点ではより一般化して記述するのが難しかったからである。必要に応じて参考にされたい。

本マニュアルは、メンバーの知見や経験にもとづく内容も多く含まれ、多様なデータベースの移行パターンを全てカバーできているわけではない。データベースレスキューは各機関やプロジェクトの予算規模、データの種類や内容、担当者によっても適切な方法は異なる。このマニュアルは、今後も適宜更新していく予定であるため、フィードバックをいただければ幸いである。それらを通じて、データベースレスキューや研究データの保存や公開、データベース設計における新たな課題の整理につながることを期待する。

### 1-3. データベースのライフサイクル

データベースレスキューの手順を考えるにあたって、最初にデータベースにもライフサイクルがあることを意識すると理解しやすい。データベースのライフサイクルはデータベースによって様々であるので一般化できないが、よく知られた「研究データのライフサイクル」を参考にしつつ、データベースのライフサイクルの一例をあげる。

1. 企画・設計：データベースを企画・設計する。資金を得る。
2. データの収集：データベースを作成するためのデータを収集する。
3. 整理・加工：設計に基づき、データを加工したりメタデータを付与するなどして整理する。
4. 構築：データベースを何らかの基盤の上に構築する。
5. 活用：データベースを活用し、研究活動を行う。
6. 公開：データベースを公開する。<sup>1</sup>
7. 維持・管理：データベースを継続して維持・管理する。場合によっては、データを追加、修正する。
8. 再利用：他の研究者がデータベースを活用し、研究活動を行う。

レスキューが必要となるデータベースは、公開や維持・管理のためのリソース（ハードウェアや人材）が確保できなくなり、再利用に問題をきたすケースといえる。また、企画・設計の際に特殊なファイル形式を採用したり、整理・加工の際に適切なメタデータを付与しなかった

---

<sup>1</sup> 科研費研究成果公開促進費の助成を受け作成されたデータベースは、公開が前提となっている。ただし、インターネットを使った公開とは限らず、CD-ROMでの提供など、公開の手段は一様でない。また、公開のタイミングもここにあげた一例のとおりのものであるとは限らない。

りした場合は、レスキューの際の困難となりうるので、このプロセスからアプローチできれば、レスキューがよりスムーズに行える。

## 2. データベースのキュレーション

この章からは、データベースのレスキューの実際について述べる。

レスキューが必要なすべてのデータベースをレスキューすることは現実的ではないため、自機関の研究データ管理ポリシーや状況に応じて、受入が可能なデータベースの条件を決めておくといよい。条件を検討するにあたって考慮すべき事項の例を挙げる。

- (1) データベースといえるか（紀要や学内刊行物、リンク集やウェブサイトでないか）
- (2) データの追加が終わっているか（これ以上増えるか）
- (3) 機関リポジトリ登録にあたってファイル形式の変換がどれくらい容易か
- (4) データベースを作成した研究者やその他の権利者からの許諾がどれくらい容易に得られるか
- (5) 既に別の形で機関リポジトリ等に登録があるか（同じデータが紀要としてまとめられている等）
- (6) 現在の運用状況はどうか（財源、維持・更新の体制）
- (7) 科学技術やインターネット技術の進歩、時間経過により該当データが陳腐化・大衆化していないか
- (8) 完全に上位互換で有名なウェブサイトやデータベースが別途存在し、独自にデータベースとして維持・管理すべきか（遺伝子情報など）
- (9) 機関リポジトリ登録にあたって難しい変換が必要でないとしても、データ自体よりもパラメータ指定や検索機能が重要なデータベースでないか（シミュレーションシステム等）
- (10) 個人情報等の機密情報や、学術や社会情勢の発展によって現在ではセンシティブな問題を引き起こす可能性のある内容を含むため、継続的な公開が不適切でないか
- (11) 学術的に重要か

## 3. レスキューの前準備

前章に基づいてレスキューを行う対象を絞ったら、実際にレスキュー作業に着手する前に、準備・確認しておくべきことについて説明する。

### 3-1. 研究者・関連部署等との交渉

- (1) データベースに関して文献調査（論文、科研費報告書等）を行い、学術的背景や成果などの情報を収集して可能な範囲で理解する。これにより、研究者とのその後のコミュニケーションがより円滑になる可能性がある。
- (2) データベース作成者やデータ権利者を特定し、コンタクトを取ったうえで、以下の内容について説明・確認する。
  - i. レスキュー趣旨の説明
  - ii. データベース作成経緯の確認
  - iii. 資金提供者との契約内容の確認
  - iv. 公開可否の確認
  - v. 他にデータの作成者・寄与者がいるかどうか（他にも確認すべき相手がいれば、同様の手順を踏む）
- (3) 作成者を含む関係者との間で、レスキュー内容やレスキュー後の運用などについて定めた覚書を交わしておくことが望ましい。覚書では、
  - i. 公開する際のライセンス
  - ii. データに法的および契約上の問題が生じた場合の対応
  - iii. ファイル形式等技術的な問題が生じた場合の対応等に触れておく和良好的。
- (4) 必要に応じて、研究・知財関連部署に、データの内容を説明したうえで、権利上の問題の有無について確認する。

### 3-2. 対象となるデータベースの現状調査と受入先の検討

#### 3-2-1. システム、ファイル形式、データ数・ファイル容量、ハードウェアの確認

- (1) ハードウェアが現在も動作するかを確認する。正しく動作しなければ、専門業者によるデータ復旧作業が必要な場合がある。
- (2) 使われているシステムやファイル形式が、一般的なもの（技術的な汎用性が高い、同分野でデファクトスタンダードである、技術情報が豊富である等）か、独自性の高いものかを確認する。後者であれば、現在でも利用するためのソフトウェアが入手可能か、セキュリティ上のリスクはないか、汎用的な形式に変換できるかを検討する。詳細が不明であれば、構築した研究者および関係者にコンタクトを取ったり、納品時のデータや構築時の仕様書を確認する。
- (3) データ数やファイル容量を確認する。レスキューの受入先で受け入れ可能かどうかの判断を行う。



- (4) 複雑なデータベースであれば、完全なレスキューは難しいこともある。技術的にレスキュー可能な部分が研究に必要不可欠な部分を満たしているかを、研究者等と協議しながら、まずは要望を満たせる最低限のレベルでレスキューすることも検討する。

### 3-2-2. 受入先のサーバ・システム

- (1) 既存のシステム（機関リポジトリ、デジタルアーカイブ等）に受け入れるか、新規システムを構築するかを検討する。後者の場合は、イニシャルコストやランニングコストが既存システムに入れるよりも大きく発生することを十分に考慮する。
- (2) 新規システムを構築する場合は、さらにサーバおよびシステムについての検討が必要である。（詳細は、付録 3. レスキュー先のハード面、4. レスキュー先のソフト面を参照）
- i. サーバをオンプレミスで構築するか、クラウドで構築するかを検討する。
  - ii. システムとしてどのようなものを採用するかを検討する。候補としては、CMS、WEKO3、各種オープンソース、静的サイトジェネレータ、図書館システム（OPAC）等がある。データベースの特性や、構築・維持体制に応じて選択する。

### 3-3. レスキュー計画の策定

データベースの容量やシステムの内容をもとに、レスキューして公開するまでのおおよそのスケジュールと分担等を決めておく。

## 4. レスキューの手順

この章では、図書館で運用している既存システム（主に機関リポジトリ）への移行を前提に一般的なレスキューの手順を記述している。新規システムへの構築については、付録 3. レスキュー先のハード面、4. レスキュー先のソフト面 を参照のこと。

基本的なデータベースレスキューの流れは、1. データの取得、2. データ本体の確認、3. メタデータの検討、4. システムへの登録、5. 公開範囲、公開日、公開方法を決定する、6. オリジナルデータベースの停止（運用中の場合）となるが、レスキューの進め方によっては手順が前後することも考えられる。処理手順、処理方法を決定する際には、以下の点を考慮するとよい。

- (1) データベースについて
- i. どのような分野を対象にしているデータベースか。
  - ii. 誰がどのように利用している（いた）のか
- (2) データ本体について

- i. 長期的な利用と保存に適した、持続可能なファイル形式は何か。

同分野でデファクトスタンダードになっている形式や、オープンで汎用的なデファクトスタンダードの形式が利用可能かどうか。

(3) メタデータについて

- i. 利用者がデータベースやデータ本体を理解するために必要な情報は何か。
- ii. 以下について十分な情報があるか 参照情報（DOI等の識別子）、文脈情報（論文や書籍等の関連資料との関係）、来歴情報（データベースの作成、データベースのレスキュー、ファイル形式の変換等）、ライセンス情報、アクセス権情報等

## 4-1. データの取得

### 4-1-1. 取得方法を決定する

レスキュー対象によるが、主に以下の手段が考えられる。

- (1) 外付けメディア
- (2) PC（サーバ）本体
- (3) メール添付
- (4) 公開中のサーバからダウンロード

### 4-1-2. ウイルススキャンの実施

システムに取り込む前に、必ずウイルススキャンを実施する。スタンドアロン環境で実施することが望ましい。

## 4-2. データの確認

### 4-2-1. データの内容を確認する

基本的には提供を受けるデータの内容は、提供側で確認済であることを前提とする。ただし、必要に応じて、データ数とメタデータのアイテム数を照合する等、取得したデータに欠損がないことを確認する。

### 4-2-2. データ容量、データファイル形式を確認する

機関リポジトリやデジタルアーカイブに保存する場合には、データ容量に問題がないか確認する。

データファイル形式について、特殊な形式や一部の企業が独占する形式が利用されているならば、オープンで汎用的な形式に変換することを検討する。同分野でデファクトスタンダードになっている形式を採用することも考慮すること。（付録 2. データの取り出し も参照）

#### 4-2-3. データベースに関するドキュメントを確認する

元のデータベースの作成に関する文書が存在すると、データの取得方法や処理方法、データベースの構築方法や利用方法等に関する情報を取得でき、レスキューする際の参考になる。データベースの来歴情報については、必要に応じて移行先のメタデータに追記する。

#### 4-2-4. データ保存の粒度を決定する

1-1 で述べた通り、データベースレスキューでは「適切にレスキューし、将来的な利用環境を提供する」ことを目的としている。その点では、原則、オリジナルのデータベースと同じ粒度でレスキュー先にデータを保存することが望ましい。しかし、3-2 で述べたとおり、検索機能を含めた「完全なレスキューは難しい」ケースもある。そのような場合は、DB の内容を一つスプレッドシートの形にまとめて登録する等、粒度を変更して登録することも考慮する。どのような粒度で登録するかは、研究者の意向や利用者の利便性を考慮すること。なお、オリジナルのメタデータと同じ粒度で登録する場合に、取得したメタデータとデータファイルの数が一致しないときは、受け取ったメディアの削除ファイル、一時ファイル等にデータファイルが存在しないか確認する。発見できなければデータベースを作成した研究者等に確認する。

#### 4-2-5. 個人情報、機密情報等の公開できない情報の有無を確認する。

公開対象のデータやメタデータに個人情報、機密情報等が含まれていないかを確認する。公開済のデータベースの公開部分は基本的には解決している可能性が高い。

公開できない情報が含まれている場合は、非公開、制限公開、部分公開、メタデータのみ公開等の処理方法を決定する。（4-4. システムへの登録、4-5. 公開範囲、公開日、公開方法を決定する を参照）

アクセス制限を設ける場合は、公開条件、公開範囲、公開期間をメタデータに記載した上で、レスキュー内容をまとめたドキュメントに理由と合わせて記録する。（4-3. メタデータの検討 を参照）

写真等が含まれる場合は、肖像権<sup>2</sup>や著作権に注意すること。研究コミュニティの慣習も確認する。

---

<sup>2</sup> 肖像権ガイドライン案 <http://digitalarchivejapan.org/bukai/legal/shozoken-guideline>

デジタルアーカイブ学会が検討中のガイドラインの素案。人物を含む写真等の肖像権を検討する際に参考になる。

## 4-3. メタデータの検討

### 4-3-1. メタデータを抽出する

MySQL 等のデータベース自体を引き取った場合、データベースからメタデータを抽出する必要がある（スプレッドシート等でメタデータを受け取った場合は不要）。抽出方法は受取ったデータベースや移行方法により異なるが、一般的には csv 形式で抽出することが多い。必要に応じてデータベースの形式や表形式での抽出方法を研究者に確認するとよい。

### 4-3-2. メタデータスキーマを決定する

分野のメタデータスキーマを採用するか、機関リポジトリのメタデータスキーマ（例えば JPCOAR スキーマ）を採用するかを決定する。登録や利用上の問題がないならば、機関リポジトリのメタデータスキーマを利用することが多い。

### 4-3-3. メタデータをマッピングする

機関リポジトリのメタデータスキーマを利用する場合、レスキュー元のメタデータからマッピングが必要になる。マッピングはメタデータスキーマの変更を伴うので、データベースを作成した研究者等の同意を得て実施すること。

マッピング作業では、オリジナルの複数のメタデータフィールドを、移行先の一つのフィールドにまとめることも考えられる。マッピングにおけるメタデータの統合はデータベースの利便性や将来の移行可能性を考慮して実施する。

また、必要に応じて、レスキュー時に新たに取得したメタデータを追加する。例えば以下の情報を追加することが考えられる。

- (1) レスキュー作業に関する情報
- (2) データベースの来歴（データ取得、データベース構築等）に関する情報
- (3) ライセンスに関する情報

これらの情報は、レスキュー内容をまとめたドキュメントとして記録する。

### 4-3-4. 必要に応じてシステムに新たなメタデータフィールドを追加する

既存のメタデータスキーマを利用しない場合は、レスキュー対象のメタデータに合わせて新たなメタデータフィールドを追加することも考えられる。新しく追加するメタデータフィールドには、テキスト、数値、日付、チェックボックス等、フィールドの型を定義するが、基本的には、オリジナルのデータベースに準じればよい。フィールドの追加はデータベースの利便性や相互運用性、維持管理のコストを考慮して実施する。

#### 4-3-5. DOI の取得

必要に応じて DOI を取得する。

junii2 ver3.1 及び JPCOAR スキーマ Ver.1.1 に対応すると、JaLC 連携で研究データに DOI を付与することが可能である。詳しくは以下を参照。

- IRDB データ提供機関のための DOI 管理・メタデータ入力ガイドライン : junii2 編  
<http://id.nii.ac.jp/1458/00000185/>
- IRDB データ提供機関のための DOI 管理・メタデータ入力ガイドライン : JPCOAR スキーマ編  
<http://id.nii.ac.jp/1458/00000135/>
- junii2 ガイドライン [https://www.nii.ac.jp/irp/archive/system/pdf/junii2guide\\_ver3.1.pdf](https://www.nii.ac.jp/irp/archive/system/pdf/junii2guide_ver3.1.pdf)
- JPCOAR スキーマガイドライン <https://schema.irdb.nii.ac.jp/ja>

#### 4-4. システムへの登録

移行するデータ数によるが、基本的にはシステムの一括登録を利用することが多い。必ず登録チェックおよびテスト登録を実施してから、本登録を実行すること。

登録数が多い場合は、システムへの負荷を考慮して、分割して実行する。

#### 4-5. 公開範囲、公開日、公開方法を決定する

##### 4-5-1. 公開範囲を決定する（制限を設ける場合）

関係者と交わした覚書や個人情報などの有無、機関リポジトリの運用ポリシーに従い公開範囲を設定する。データに個人情報等の法的、倫理的問題や、データ取得の際の契約上の問題が含まれている場合は、公開範囲の設定に注意すること。必要に応じて研究者に相談する。公開範囲の制限は以下のパターンが考えられる。

- (1) 非公開：データベース自体を公開しない
- (2) 制限付き公開<sup>3</sup>：学内限定等、公開範囲を制限して公開する
- (3) 部分公開：該当するデータのみを非公開にする
- (4) メタデータのみ公開：メタデータは公開するがデータ本体は非公開とする

##### 4-5-2. 公開するデータのライセンスを決定する

自由利用できるライセンスで公開して、メタデータに明記することを推奨する。例えば以下のライセンスの利用が考えられる。

- (1) CC0 <https://creativecommons.jp/sciencecommons/cc0-faq/>
- (2) CC BY 4.0 <https://creativecommons.org/licenses/by/4.0/deed.ja>

公開ライセンスは、データ取得の段階で研究者との間で合意を得ることが望ましい。

---

<sup>3</sup> 制限付き公開についてはシステムで対応可能か事前に確認すること。

### 4-5-3. 公開方法

機関リポジトリ（移行先のシステム）の仕様に従う。

特殊な閲覧システムを用いたシステムをレスキューする必要がある場合は、閲覧方法に制限があることを注記する。

### 4-5-4. 公開日の決定

必要に応じて、研究者と相談して公開日を決定する。公開中のデータベースをレスキューする場合は、切り替え時期を考慮して決定する。

## 4-6. オリジナルデータベースの停止

オリジナルのデータベースは、レスキューしたデータベースを公開した後に停止する。必要に応じて、オリジナルデータベースの URL からレスキュー先へリダイレクトを設定することが望ましい。

## 5. レスキューしたデータベースの維持管理

レスキューしたデータベースの維持管理は、機関リポジトリ（移行先のデータベース）の運用規定及び研究者との間の覚書（3-1. 研究者・関連部署との交渉 を参照）に従う。

### 5-1. 継続的なメンテナンス

#### 5-1-1. 公開データのバックアップ

取得時のデータについては、外部メディア等に保存しておくこと。複数のメディアに保存することを推奨する。

登録したデータのバックアップは、機関リポジトリ（移行先システム）定期バックアップに含まれる。（付録 5. バックアップ体制 も参照）

#### 5-1-2. データ形式の変換作業（マイグレーション）

レスキューしたデータベースのデータが特殊なファイル形式である場合は、ファイル形式の変換作業等が必要になる。変換作業はファイル形式が陳腐化する度に必要になる。データベースの価値やアクセスログ、作業に要する労力や経費等から継続的に変換作業を実施するかを判断する。（付録 2. データの取り出し を参照）

## 5-2. ドキュメントの整備

レスキューの経緯、レスキューの内容等のドキュメントを整備する。システムリプレイスやデータ形式の陳腐の結果、将来的にデータやデータベースの移行作業が発生する可能性がある。その際には過去に行った移行作業の記録が参考になる。レスキューを実施した際には、必ずドキュメントを整備し保存すること。オリジナルのデータベースにドキュメントが存在するならば、それを参考に必要に応じて補足する。

ドキュメントは以下に留意してまとめると良い。

(1) なぜレスキューする必要が生じたのか。

例) 管理者の退職、ハードウェアやソフトウェアの老朽化等。

(2) どのようにレスキューを実施したか。

(3) 採用したレスキュー方法が、将来的に維持・管理の問題となる可能性はないか。

## 5-3. 利用統計のフィードバック

コンテンツの利用状況は定期的に研究者に通知する。システムの自動通知機能を利用することもできるだろう。そのようなシステムを有していない場合は、通知する手段を検討する。利用が多いことが研究者に周知されれば、今後、研究者がデータベースの構築を検討する際に機関リポジトリの活用が候補に入ることも期待できる。

## 5-4. データの追加・修正

### 5-4-1. データ本体の追加

データ追加の要望が研究者から寄せられた場合、データ数、データ容量、ファイル形式等の確認を行う。システムのスペックや作業内容等を考慮した上で、追加が可能であればメタデータの作成を依頼し、登録する。

### 5-4-2. メタデータの追加・修正

誤字脱字レベルの修正については、機関リポジトリ担当者で実施する。

メタデータの内容に関する修正は研究者に相談する。

## 5-5. 連絡体制の構築

内容についての問い合わせがあったときの対応等を考慮し、研究者との連絡体制を構築・維持すること。

## 6. データベースの取下げ・消去

ここでは、一度レスキューしたデータベースや一部のデータを、取下げなければならない事情が生じた場合に、取るべき基本的な対応について説明する。

### 6-1. 取下げ・消去の検討が必要になるケース

(1) 作成者その他のステークホルダーから取下げの要求があった

事前に取り交わした覚書の中に、取下げに関しての規定があれば、それに該当するかどうかを確認する。それに該当しない、あるいは事前に取り下げについて定めていなかった場合は、要求の趣旨をよく確認し慎重に対応する。

(2) 法的・倫理的な問題が指摘された

次のようなケースが考えられる。いずれの場合も、指摘の内容を精査したうえで、慎重な対応が求められる。

- i. 第三者の知的財産権（著作権等）を侵害している
- ii. 第三者のプライバシーやセンシティブ情報に触れている
- iii. その他、法的・倫理的に問題があるデータを含む

(3) その他

それ以外のさまざまな理由で取下げの検討が必要となった場合は、事前の覚書に即して、作成者等と協議を行う。安易な取下げにならないよう、慎重な検討が求められる。

### 6-2. 取下げができないケース

DOIの取得後は、原則として取下げは行うべきではない。関係者と協議のうえ、どうしても取下げざるを得ない場合は、DOI登録機関（JaLC、Crossref等）の指示に従い、取下げた旨と理由を表示するランディングページを維持するなどの対応を取る。

### 6-3. 取下げの手順

以下は機関リポジトリで公開していたデータを取下げる状況を想定している。その他のシステムで公開していた場合は、それぞれのシステムに合わせて読み替える。

- (1) データを非表示・アクセス不可とする
- (2) 公開していたURLまたはそのリダイレクト先に、公開を取下げた旨の表示を行う
- (3) データのバックアップを取る
- (4) 公開サーバからデータ本体を削除する（メタデータのページは可能な限り維持する）
- (5) 関係者に取下げ報告を行う
- (6) データ公開時に何らかの広報を行っていれば、同じルートで取下げの周知を行う



## 6-4. 取下げた後の処理

### 6-4-1. 再公開の可能性を検討

データを保持しておく必要があるかを決めるために、再度公開の可能性があるかどうか、作成者の意向や取下げの理由をもとに検討する。

### 6-4-2. バックアップを取ったデータの扱い

再公開の可能性があれば、バックアップデータを保管する。あわせて、取下げに至った経緯も含めて、担当者間で共有・引き継ぎを行い、再公開に備える。

再公開の可能性がなければ、データを破棄したり、関係者のみ利用可能とするなど扱いを検討する。この場合も、将来の同様のケースへの対処の参考となるように経緯について記録しておく。

## 7. データベースの持続可能性向上のための研究者向けアドボカシー

この章では、貴重な研究成果としてのデータベースを持続的に継承するために、機関リポジトリやデジタルアーカイブの担当者がデータベースを作成しようとする研究者にどのような情報提供を行うとよいかについて述べる。前章までに述べたレスキューの作業や手続きで手間どったり、そもそもデータベースの維持・管理が難しい状態に陥ったりするのは研究者にとっても避けたいことである。研究者によるデータベースの持続可能性や、いざというときのレスキューのしやすさを向上させるための考え方についてアドボカシーを日常的に行っておくことが重要である。アドボカシーの方法としては、研究データ管理に関するガイドラインに以下の内容を盛り込んだり、セミナーやウェブサイトを通じてのファカルティ・ディベロップメント、データベース構築時の相談対応や DMP の作成支援などが考えられる。機関リポジトリやデジタルアーカイブの担当部門だけでなく、URA など研究支援・推進を専門とするスタッフとの協働も効果的だろう。

特に、どのような考え方で用いる技術を選定するかや、誰がどのようにシステムを構築・維持するかは、データベースの企画時点から十分に検討しておくよう助言できるとよい。

まず、システムを選定する際には汎用的でグローバルに人気があり、積極的な開発や情報交換が行われているソフトウェアやサービスを選定するようにするとよい。データベースを運用中に、そのソフトウェアやサービスの提供が終了するとシステムを維持できなくなるためである。もし構築や維持を行うために十分なスキルを持った担当者がいない場合には、業者に開発と保守を発注することも考えられるが、その際にも可能な限り独自開発システムを避け、広く

普及している CMS やオープンソースを利用して開発したほうが、ベンダーロックの可能性が下がり、メンテナンス性も高まる。

データベースを構築して持続的に運用するには、維持・管理の人材の確保や体制づくりも考えておく必要がある。データベースを支えるシステムは、セキュリティアップデート等の保守やデータのメンテナンスも適切に行わなければならない。自前で保守を行う場合には、現時点での担当者が構築や維持を行っても、その担当者が異動してした際に維持できるかに留意する。定期的にデータを追加していくデータベースの場合は、その体制づくりも必要になる。

また、レスキューの手段として、データベースをシステムごと残す方法だけでなく、データを機関リポジトリ等へ移す選択肢があることも、同時に周知すると良い。動的なデータベースのままでは難しくても、静的なデータとしての保存であれば、かなりのデータ量であっても可能となってきた。

## 8. 今後の課題

最後に、現時点では検討が不十分で本マニュアルには詳細に記載できなかった点について触れておく。将来的には、レスキュー事例や研究データ管理の経験が蓄積したのちに、これらの課題についてあらためて検討したい。

### 8-1. メンテナンスコストと持続可能性

メンテナンスにかかるコストを考慮し、ファイルを圧縮して公開する等、維持管理方法を変更する場合もありうる。どのデータの維持管理方法を見直すかは、利用状況を定期的に分析し、利用が少ないものとするなど、選定条件は慎重に検討する必要がある。

### 8-2. ライセンスの考え方

オープンサイエンスの文脈では、FAIR 原則に則って、なるべくオープンなライセンスでデータを公開することが求められる。機関リポジトリやデジタルアーカイブでデータベースを公開する場合も、当然ながらそのようなライセンスが望ましいが、研究者がデータベースを作成する際には、再利用が難しい条件を付していたり、他者が権利を持つコンテンツについての利用についての契約が曖昧であったりするため、レスキューを行う際に困難が生じる場合もある。研究者がデータベースを作成する際に、ライセンスをどのようにすればデータベースの持続可能性につながるのかを検討し、研究者に情報提供する必要がある。

## 参考文献

Cooper T. Clarke, Hilary Szu Yin Shiue. [Data Rescue Processing Guide: A Practical Guide to Processing Preservation-Ready Data from Research Data Collections. Version 1.0.](#) 2020. <https://doi.org/10.13016/dif5-arr2>

Robert R. Downs, Robert S. Chen, "[Curation of Scientific Data at Risk of Loss: Data Rescue and Dissemination](#)," in [Curating Research Data. Volume One: Practical Strategies for Your Digital Repository](#), ed. Lisa R. Johnston (ACRL, 2017). <https://doi.org/10.7916/D8W09BMQ>

National Digital Stewardship Alliance (NDSA). [Curatorial Guidance with Decision Guide and Decision Tree](#). 2020. <https://doi.org/10.17605/OSF.IO/Q8C47>

Research Data Alliance. [Guidelines for Data Rescue](#). 2017.

Andrea K. Thomer, Nicholas M. Weber, Michael B. Twidale. "[Supporting the Long-term Curation and Migration of Natural History Museum Collections Databases](#)." *Proceedings of the Association for Information Science and Technology* 55(1) (2018): 504-513. <https://doi.org/10.1002/pra2.2018.14505501055>

Andrea K. Thomer, Alexandria Jane Rayburn, Allison R. B. Tyler. "[Three Approaches to Documenting Database Migrations](#)". *International Journal of Digital Curation* 15(1) (2020): 1-5. <https://doi.org/10.2218/ijdc.v15i1.726>

三上絢子. [令和元年度実務研修報告書](#). 2020.

JPCOAR 研究データ作業部会データベースレスキュープロジェクト. [データベースレスキュープロジェクト：2019 年度の活動とレスキュー事例](#). 2019. <https://jpcoar.repo.nii.ac.jp/records/237>

西蘭由依, 中筋知恵, 芝翔太郎, 小林裕太, 松野渉, 常川真央, 小野寺千栄, 南山泰之. [教材「研究者のための研究データマネジメント」](#). 2020. <https://jpcoar.repo.nii.ac.jp/records/294>

吉田幸苗, 天野絵里子, 松本侑子, 西蘭由依, 山地一禎, 南山泰之, 尾城孝一, 常川真央, 大園隼彦. [教材「研究データ管理サービスの設計と実践」第 2 版](#). 2021. <https://jpcoar.repo.nii.ac.jp/records/607>

# 付録：システムを含めてデータベースをレスキューする場合の技術的な課題

## 1. はじめに

本作業部会ではデータベースレスキューにおいて、機関リポジトリやデジタルアーカイブなどの既存の基盤ではなく、元の検索機能なども含めたシステムごと移行する場合における技術的な課題について付録資料としてまとめた。

本資料は以下の4項目ごとに内容を分類し、課題と解決案を列挙する。

- データの取り出し
- レスキュー先のハード面
- レスキュー先のソフト面
- バックアップ体制

## 2. データの取り出し

### 2-1. 現状の公開システムおよびデータファイル形式の把握

データベースレスキューのためには、まず現状の公開システムについて把握する必要があるが、レスキューすべきデータベースでは、研究者が独自に構築したものや、業者に依頼し独自構築されたものが多い。レスキュー後の公開の継続性を考えると、汎用的な公開方法を検討すべきであり、基本的にはシステム以外の公開データおよびサイトの情報のみをレスキューし、公開システムは別途検討することが望ましい。

### 2-2. レスキューすべきデータを確認する

- (1) 構成が不明であれば、構築した研究者及び関係者にコンタクトをとったり、業者納品時のデータや構築時の仕様書を確認したりする。
- (2) システムのバージョンが古かったり、現在では利用できないファイル形式を使用している現在のPC等で確認できない場合は、該当するシステムやファイル形式を開くことのできる互換性のあるファイルソフトが提供されていないか、該当するシステムやファイル形式を開くことのできる端末が自機関にないか確認する。その際は、古いシステムやファイル形式を利用することによるセキュリティ上のリスクがないか十分検討する。  
(オフラインで作業する等)
- (3) ハードウェアが古く適切に動作しない場合は、業者へのデータ復旧作業の依頼も検討する。

- (4) 複雑なサイトの場合、すべてのレスキューは難しい場合がある。技術的にレスキュー可能な部分が研究に必要不可欠な部分を満たしているかデータ提供元と協議しながら、まずは要望を満たせる最低限のレベルでレスキューすることも検討する。

## 2-3. データの抽出

- (1) レスキュー先に移行する際に必要な画像データや文書データ、html ファイルなどをレスキューする。
- (2) もし元のデータベースのシステムが移行先のデータベースと互換性があれば、丸ごと移行できる場合もある。
- (3) 該当するファイルが見つからない場合は、元のシステムから汎用的な形式にデータが抽出できないか確認する。
- (4) 汎用的な形式とはオープン化（一般的なツールで再生できる、一部の企業が独占する技術ではない）と使用実績という点で評価できる。以下に参考資料を示す。
- a. 国立国会図書館が採用するファイルフォーマット  
<https://www.ndl.go.jp/jp/preservation/dlib/formats.html>
  - b. 国立国会図書館資料デジタル化の手引  
<https://www.ndl.go.jp/jp/preservation/digitization/guide.html>
  - c. デジタルアーカイブのための長期保存ガイドライン（2020 年版）（デジタルアーカイブジャパン推進委員会及び実務者検討委員会）  
[https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive\\_suisiniinkai/index.html](https://www.kantei.go.jp/jp/singi/titeki2/digitalarchive_suisiniinkai/index.html)
  - d. 古典籍のデジタル化（撮影）マニュアル（国文学研究資料館）  
<https://www.nijl.ac.jp/pages/cijproject/database.html#section03>
- (5) 抽出できない場合、もしまだウェブ公開されている場合はスクレイピングしてデータ取得することもできる。ウェブ公開していない場合でも、ローカル上でサイトの公開が再現できれば同様にスクレイピングでの取得も可能と考えられる。
- (6) 抽出できない場合、かつウェブ公開されていない場合は、業者に依頼もしくは有料もしくは無償ツールを利用して自機関で抽出する。

## 3. レスキュー先のハード面

レスキュー先のシステムを選定する際には、ハード面・ソフト面を考慮する必要がある。以下ではまずハード面で考慮することについて述べる。

最初に検討するのは、既存のシステムにレスキューするか、新規にシステムを構築するかである。

### 3-1. 既存のシステムにレスキューする場合

- (1) 移行可能なファイル形式、ファイルの合計サイズが容量上問題ないかを確認する。
- (2) 現時点では容量上問題ない場合でも、今後データが増加する可能性がないか検討しておく。

### 3-2 新規にシステムを構築する場合

新規にシステムを構築する場合、オンプレミスのサーバに構築するか、クラウド上に構築するかを選択がある。

- (1) 自前のサーバを用意できる場合であっても、OS のアップデートや公開に必要なソフトウェア等の対応が自らで行えるか十分に検討しておく必要がある。もし、特定のスキルのある人員に頼った構築や運営をしてしまうと、その人材が退職や移動した場合に維持できなくなり、レスキュー元のデータベースと同じ状況になりかねない。スキルのある人材がいない場合は、業者への保守依頼も検討する。保守にはサーバ自体の保守とシステムの保守で複数の契約が必要となる場合もある。また、基本的には 5 年程度でサーバの耐用年数が来るため、定期的にサーバの更新費用も必要となる。なお、最近はサーバを自前で持つことを禁止している機関も多いため自機関のセキュリティポリシー等を事前に確認しておく必要がある。
- (2) クラウド上に構築する場合は、サーバの購入や更新費用は掛からないが、毎年一定の利用料を支払う必要がある。クラウドの選択肢としては、各機関の機関リポジトリは有力な検討候補となる。NII と JPCOAR で開発を進めている WEKO3 は簡易的なデジタルアーカイブの機能も提供される予定である。また、Internet Archive 等海外等のデータアーカイブサイトの検討も有用である。

## 4. レスキュー先のソフト面

特に新規でシステムを構築する場合には、以下の点に留意する必要がある。

### 4-1. どのようにシステムを構築するか

レスキュー対象となっているデータベースには、業者や研究者が開発を行った後メンテナンスが行き届かず維持できなくなったものがある。基本的には同じ轍を踏まないよう、メンテナンスの容易さ、継続可能性を重視すべきであるが、移行後のデータベースが従来のデータベースと比較してあまりに不便になることがないよう、バランスをとって検討する必要がある。以下は選択肢の例である。



### (1) CMS(Contents Management System：コンテンツ・マネジメント・システム)

CMSとはウェブサイトのコンテンツを構成するテキストや画像、デザインなどの情報を統合・体系的に管理し、配信など必要な処理を行うシステムのことである。CMSには、一般的なサイトにも広く利用されているWordPressや、東京大学学術資産等アーカイブズ共用サーバ (<https://iif.dl.itc.u-tokyo.ac.jp/repo/>)で使用されているOmeka S、京都大学貴重資料デジタルアーカイブ (<https://rmda.kulib.kyoto-u.ac.jp/>)で使用されているDrupalなどがある。CMSは、CMSを入れるだけである程度のサイトが作れてしまう、専門知識や技術的な知識がなくてもコンテンツを用意すればウェブページの更新・追加が可能である、プラグインやモジュールの追加で拡張性が高い、などのメリットがある。一方で、プラグインやモジュールを追加するほど、本体のアップデートや更新が行いにくくなる場合が多く、管理できる者がいない場合メンテナンスが行き届かなくなる可能性がある。

### (2) WEKO3（次期 JAIRO Cloud）

現在国内の機関リポジトリシステムとして最も利用されているWEKO2の後継システムとして開発が進んでいるWEKO3では、研究データの公開に対応した機能が提供される予定である。また、WEKO3ではマルチリポジトリ機能というインデックスツリーを別のサイトとして公開できる機能が提供される予定であり、複数のデータベースを一つのサイトで公開することができる。また、IIIF画像の表示などデジタルアーカイブに関する機能も提供される予定であり、デジタルアーカイブサイトの移行も検討できる可能性がある。サイトの構築やメンテナンスについてあまり考慮する必要がないため人的資源がない場合には非常に有用な選択肢となる可能性があるが、WEKO3の利用にはストレージ使用量に応じた JAIRO Cloud 特別料金が発生するため、データの容量によっては維持費用がかさむ可能性がある。

### (3) オープンソースの活用

サイトを構築する際に、オープンソースをいくつか組み合わせてシステムを構築することで、コストが節約できたり、システムのメンテナンス性が向上したりする可能性がある。青森県立図書館デジタルアーカイブの取り組みが非常に参考となる。

[https://jpsearch.go.jp/static/pdf/event/libraryfair2020/lff2020\\_forum4.pdf](https://jpsearch.go.jp/static/pdf/event/libraryfair2020/lff2020_forum4.pdf)

### (4) 静的サイトジェネレータの利用

静的サイトジェネレータとは、HTMLのような静的コンテンツだけでサイトを作成するサービスのことであり、CMSよりもセキュリティ上の懸念が少なくメンテナンスが容易、サイトの表示速度が速いなどのメリットがある。ただし、頻繁にデータ更新があったり、コンテンツ量が膨大になったりするサイトの作成にはあまり向いていない。人気のある静的サイトジェネレータについては以下のサイトで確認ができる。

<https://jamstack.org/generators/>

(5) 国内外で実績のある業者のサービスを利用する

もし十分な構築費および保守費を確保できるのであれば、国内外で実績のある業者のサービスを利用することも十分な選択肢である。

(6) 国内外のリポジトリ・アーカイブサイトを利用する

自前での構築が困難な場合、国内外の研究データや汎用データのリポジトリ、アーカイブサイトなどにデータを保存して公開することも選択肢としてあげられる。データベースのレスキューではないが、東京大学のデジタルアーカイブで、Internet Archive というアメリカのアーカイブサイトと連携し、画像とメタデータを取り込んでもらい公開を行っている例がある（<https://archive.org/details/utokyo-general>）。無償で利用できる場合もあるが、機関利用の場合やデータ容量によっては利用料がかかる場合もある。

(7) 蔵書検索サーバ（Web-OPAC）を利用する

図書館システムのリース契約に含まれている OPAC サーバを使用し、通常の蔵書検索と同じ検索窓からレスキュー資料を検索し PDF を表示する方法もある。この場合、サーバ構築・維持の費用はリース料金に含まれ、業者による保守の対象となる。ただ、図書館システム更新のたびにデータ移行が必要となりその都度構築費用がかかるというデメリットもある。一例として、緑丘アーカイブズ（小樽商科大学）がある。

[https://webopac.ih.otaru-uc.ac.jp/index.php?action=pages\\_view\\_main&active\\_action=v3search\\_view\\_main Inklst &block\\_id=296&tab\\_num=0#v3search\\_view\\_main Inklst](https://webopac.ih.otaru-uc.ac.jp/index.php?action=pages_view_main&active_action=v3search_view_main Inklst &block_id=296&tab_num=0#v3search_view_main Inklst)

## 4-2. 構築にあたって留意すること

構築にあたっては、レスキュー元のサービスを再現することももちろん大事であるが、そのために特別な開発を行ったり CMS で多くのプラグインやモジュールを追加したりすると、アップデートが困難になったりメンテナンスが行き届かなくなったりするため、可能な限り既存の仕組みや既存のオープンソースの仕組みで再現できないか検討し、場合によっては割り切って不要な機能をそぎ落とすことも大事である。

## 5. バックアップ体制

### 5-1. データの責任の所在を明らかにする

レスキューしたデータがサーバの故障等で消失してしまう可能性があることを考えておかなければならない。レスキューしたデータをデータ提供元、公開元のどちらがバックアップしておくかについて、公開前にあらかじめ取り決めておくべきである。



## 5-2. バックアップ方法

レスキューした側でデータの責任を持つとした場合、バックアップ体制を明確にしておく必要がある。

### 5-2-1. バックアップする対象

システム全体か、公開データのみとするかを決定する。

### 5-2-2. バックアップ手段

- (1) 公開データおよびシステムデータ全体をバックアップする場合、NAS や別のサーバにバックアップする。
- (2) 公開データの場合、別サーバ、NAS、テープドライブ、光学ディスク（DVD or Blue-Ray or CD）、HDD or SSD、クラウド上などバックアップ頻度、データの増加の可能性の有無、簡便さ、容量のバランスを比較し、適切なものを選択しバックアップを行う。
- (3) 国内外の研究データ向けの汎用リポジトリやアーカイブサイトにデータを保存しておくこともバックアップ手段の一つとして考えられる。

### 5-2-3. バックアップの体制

可能であれば一つの手段だけに頼らず、複数のバックアップを持っておくほうが望ましい。また、データの増加の可能性がある場合には、バックアップの頻度も検討する必要がある。例えば、利便性と災害対策を考えれば、バックアップの行いやすい NAS および遠隔地での保管が可能なテープやディスク形式で二重化、三重化しておくといよい。