

IRDB-カラクリと役割

どこから・どこへ・どのように

株式会社メタ・インフォ

<https://www.meta-info.co.jp>

井村 邦博



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-sa/4.0/deed.ja>

発表内容(50分)

1. 自己紹介
2. IRDBとは
3. JPCOARとIRDB
4. どこから (IRDBへのデータ収集先とプロトコル)
5. どこへ (IRDBから外部へのデータ提供)
6. どのように (IRDBの動作)
7. まとめ

自己紹介

会社名:株式会社メタ・インフォ 井村 邦博

所在地:東京都港区虎ノ門、山口県岩国市

事業内容:Drupalを利用したシステム開発、Webサイト構築

国内でのIIIF対応のデジタルアーカイブ構築や

図書館Webサイトの構築の実績多数

IRDBとの関わり:2017年のリニューアルのため、新規開発から現在まで、

開発・運用に携わる

IRDBとは

IRDBとは

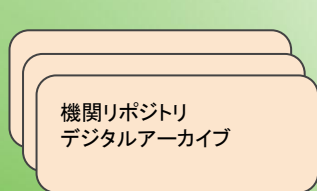
IRDB(学術機関リポジトリデータベース : Institutional Repositories DataBase)は、日本国内の学術機関リポジトリに登録されたコンテンツの**メタデータ**を収集し、提供するデータベース・サービスです。

収集したメタデータについて大きく次の機能があります。

- ① 収集したメタデータを一括で検索
- ② 収集したメタデータをチェック行い外部に提供
(CiNii、国立国会図書館、ディスカバリサービスなど)
- ③ DOIの有効化(JaLCとの連携)
- ④ 月次での統計情報の作成(全体、リポジトリ毎)

参考ページ:IRDBのサポートページ(<https://support.irdb.nii.ac.jp/ja>)

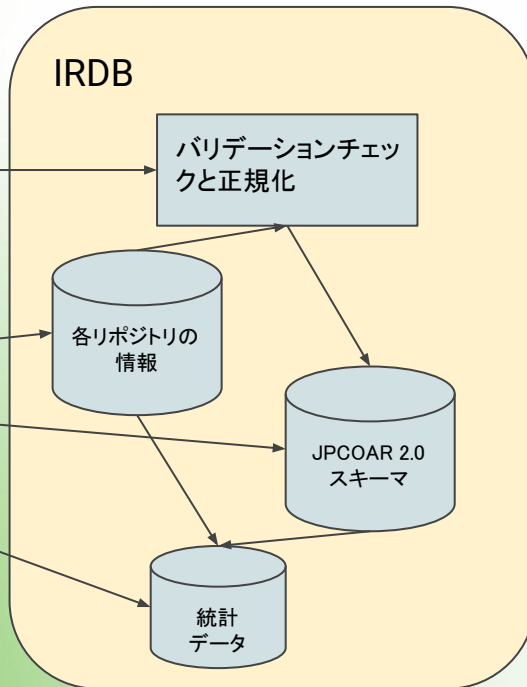
IRDBの構成図



メタデータ形式
・JPCOAR
・JPCOAR 1.1
・JPCOAR 2.0
・Junii2



ご担当者様



ジャパンリンクセンター(JaLC)

DOI有効化

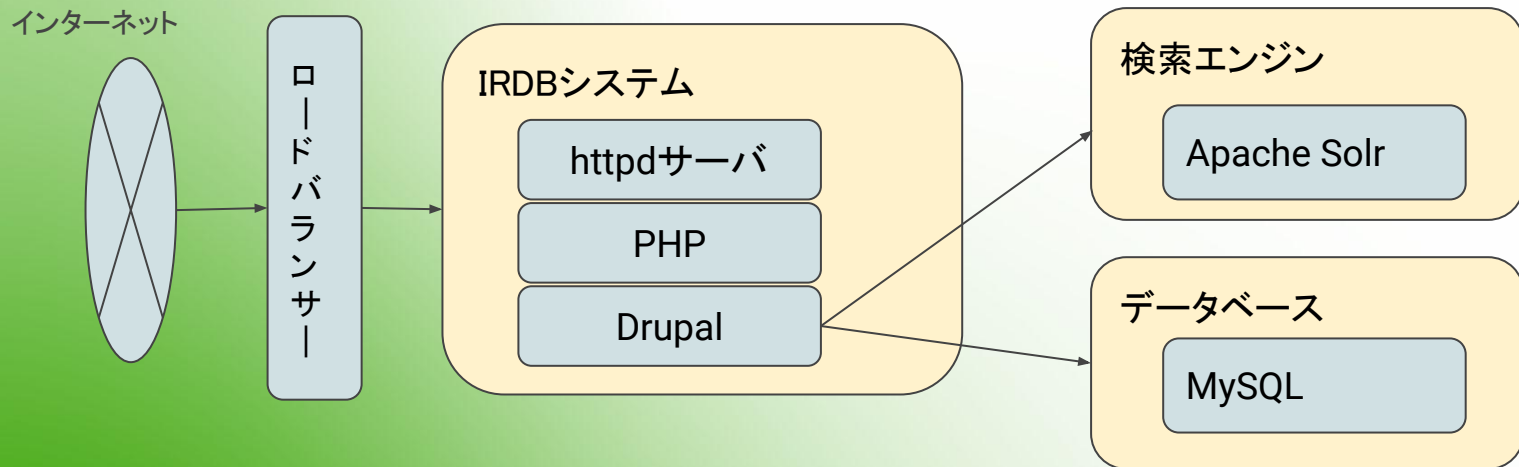
CiNii Research

国立国会図書館

医学中央雑誌刊行会

OpenAIRE など

IRDBのシステム構成図



- ・すべてのソフトウェアはオープンソースソフトウェア(OSS)を利用しています。
- ・IRDB開発当初から6年経過していますが、各ソフトウェアはバージョンアップを重ねて、サポートが切れないうに対応しています。
- ・OSSを利用することで、セキュリティを担保した上でシステムの持続性・保守性が高いと考えています。

JPCOARとIRDB

IRDBは、junii2とJPCOAR(1.0、1.0.2、2.0)のスキーマについて収集をしています。現在はJAIRO CloudがWEKO3となり、JPCOAR 1.0のスキーマで収集が増えました。

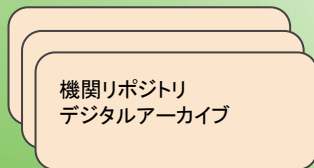
IRDBで収集したスキーマは、バリデーションチェックや正規化を実施して、内部的にはJPCOAR 2.0に変換して保存しています。最新のJPCOARスキーマバージョンに対応したシステムとなっています。

参考ページ: JPCOARスキーマガイドライン(<https://schema.irdb.nii.ac.jp/ja>)

どこから・どこへ・どのように

どこから

どこから(データを収集)



メタデータ形式

- JPCOAR
- JPCOAR 1.1
- JPCOAR 2.0
- Junii2

各リポジトリシステムから、毎日データを収集。
(水曜日は予備日)

曜日単位で、各リポジトリからデータの差分を収集しています。

各リポジトリは1週間単位でハーベストされている状況です。

IRDB

バリデーションチェックと正規化

各リポジトリの
情報

JPCOAR 2.0
スキーマ

どこから(IRDBへのデータ収集プロトコル)

IRDBは、データ収集するためのプロトコルは以下の2つに対応しています。(プロトコルは、コンピュータでデータをやりとりするために定められた手順や規約)

1. OAI-PMH

メタデータ交換を行うためのプロトコル。2001年に仕様策定。2002年にVer2の仕様策定。

データを提供する側(データプロバイダと呼ぶ)

データを収集する側(ハーベスターと呼ぶ) データ収集を「ハーベストする」

2. ResourceSync

ウェブサーバの間で**コンテンツの同期**を行うためのプロトコル。2017年に仕様策定。

コンテンツ1件毎(アイテム1件)のURLと更新日付を元にやり取りします。

追加仕様として、データを受信する側が更新通知を受け取るための仕様もあります。

どこから(IRDBへのデータ収集プロトコル)

OAI-PMHは早くに仕様策定されたこともあり、OAI-PMHが利用されていることが多いです。JAIRO Cloud(Weko3)や国立国会図書館では、OAI-PMHが利用されています。

OAI-PMHは、1回での取得件数があり(システムによって50, 100, 1000など)、その1回の取得中でエラーが発生するとデータが特定しにくいいため、リトライ処理の実装が難しいです。

ResouceSyncは、1件毎にデータ取得するためエラーデータが特定しやすく、リトライ処理が容易になります。

どこから(ID付与)

各リポジトリシステムとIRDBのシステム間でのデータやり取りに欠かせないのが、アイテムのIDとなります。(OAI-PMHでは、OAI-IDと呼ぶ)

このアイテムのIDをもとに、IRDBは新規に登録するデータなのか、更新すべきデータなのかを判断して、データ収集を行っています。

そして、IRDBにデータ登録後はIRDBの内部では全体を統一した形でIDを新たに付与しています。

どこから(ID付与)

IRDB側でデータ登録後は、ユニークなIDが付与されて管理されます。このIDがCiNiiや国立国会図書館など、世界中に流通するIDとなります。

■ IDの例

タイトル:2023年度 第22回JPCOAR運営委員会 議事次第

URL: <https://irdb.nii.ac.jp/03873/0005964451>

OAI-PMHのID: oai:irdb.nii.ac.jp:03873:0005964451

ResourceSyncのID: https://irdb.nii.ac.jp/03873/0005964451

どこから(収集したデータのIDの管理)

IRDBではIDを付与していますが、各機関リポジトリ側でもIDが付与されています。

(各機関リポジトリのIDの例)

タイトル:2023年度 第22回JPCOAR運営委員会 議事次第

OAI-PMHのID:oai:jpcoar.repo.nii.ac.jp:02000339

IRDBでは各機関リポジトリ側でのIDとIRDBで付与したIDを一对一にマッピングして、管理しています。このIDの管理によりデータの更新・削除ができます。

oai:jpcoar.repo.nii.ac.jp:02000339 → oai:irdb.nii.ac.jp:03873:0005964451

* ResourceSyncの場合は、アイテムのURLがIDとなります。

どこへ

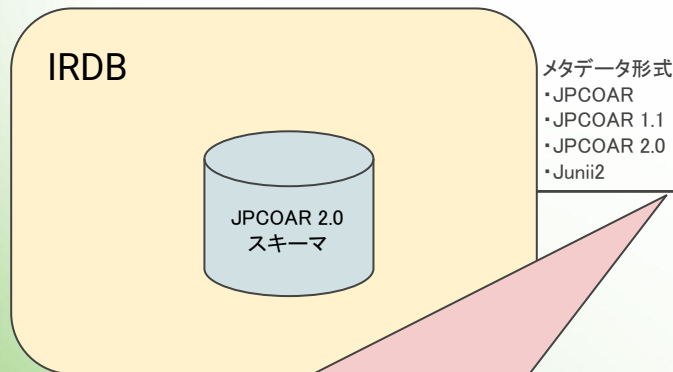
どこへ(IRDBから外部へのデータ提供)

IRDBへのデータ収集のときに、データ収集のためのプロトコル2つを説明しましたが、外部へのデータ提供についても、IRDBでは2つのプロトコルに対応しています。(OAI-PMHとResourceSync)

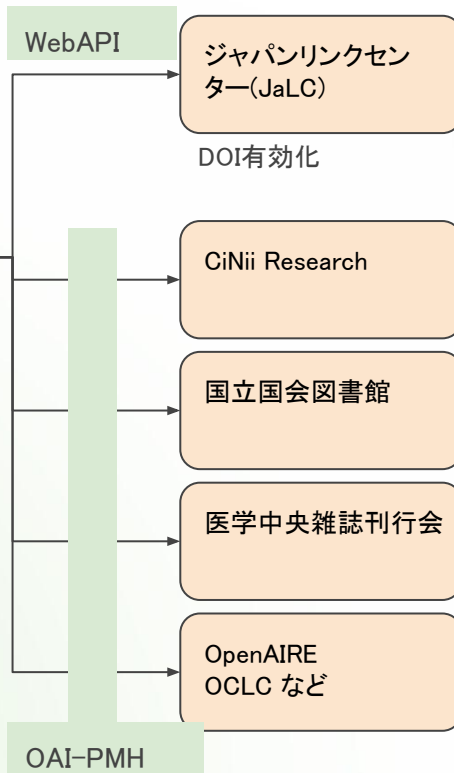
なお、実際の運用的にはOAI-PMHで提供されている形で、ResourceSyncでの提供はまだ実運用はないです。

このためOAI-PMHのIDが、世界中に流通されているIDとなります。

どこへ(IRDBから外部へのデータ提供)



JaLCを除く外部のサービスからは、任意のタイミングでIRDBからデータを取得しています。
新規作成・更新があったデータを各サービスは定期的に収集しています。
JaLCについては、毎日IRDBからWebAPIでDOIの有効化処理のためにデータを登録しています。



どこへ(JaLC以外との連携)

CiNii、国立国会図書館や各種ディスカバリーサービスとの連携は、すべてOAI-PMHの処理で取得されています。各サービス側が、任意のタイミングでIRDB側にOAI-PMHのプロトコルを利用してデータを取得しています。

このデータ取得のときにIRDB側で付与してIDが、各外部サービス側で管理される形になります。

結果、IRDB側で付与したIDをもとに各サービスはデータの更新、削除を実施します。

どこへ(JaLC連携)

IRDBの大きな機能の一つにDOI(永続的な識別子)の有効化があります。

各リポジトリで付与されたDOIについて、JaLCに登録して有効化されています。

JaLCに登録する処理は、以前はOAI-PMHでJaLCからIRDBのデータを収集していました。

しかし、1年ほど前からIRDB側からWebAPIを利用して、JaLCに登録(デポジット)を行う形に変更しました。IRDB側がDOIの有効化処理のタイミングをコントロールする形になり、エラー処理の対応も実施します。

なお、DOIは、JaLCのDOIの他にCrossrefとDataciteのDOIについてもJaLCにデポジットして有効化されます。

どのように

どのように(IRDBの動作)

あらためてIRDBの機能についての一覧になります。

- ① 収集したメタデータを一括で検索
- ② 収集したメタデータのチェックを行い、外部に提供
(CiNii、国立国会図書館、ディスカバリーサービスなど)
- ③ DOIの有効化(JaLCCとの連携)
- ④ 月次での統計情報の作成(全体、リポジトリ毎)

どのように(収集)

- ② 収集したメタデータのチェックを行い外部に提供
(CiNii、国立国会図書館、ディスカバリーサービスなど)

どのように(収集)

タイトル

「2023年度 第22回
JPCOAR運営委員会 議事次第」

のデータで具体的な処理を説明します。

アイテムURL : <https://jpcoar.repo.nii.ac.jp/records/2000339>

アイテム

運営委員会

2023年度 第22回JPCOAR運営委員会 議事次第

<https://doi.org/10.34477/0002000339>

名前 / ファイル	ライセンス	アクション
 第22回 JPCOAR運営委員会議事次第.pdf (247 KB)		ダウンロード Information

Item type	デフォルトアイテムタイプ (フル) (1)
公開日	2023-11-21
タイトル	
言語	ja
タイトル	2023年度 第22回JPCOAR運営委員会 議事次第
出版者	
言語	ja
出版者	オープンアクセスリポジトリ推進協会
日付	
日付	2023-11-21
日付タイプ	Issued
言語	
言語	jpn
資源タイプ	
資源タイプ識別子	http://purl.org/coar/resource_type/c_1843
資源タイプ	other
ID登録	
ID登録	10.34477/0002000339
ID登録タイプ	JaLC

[戻る](#)

84 views
total
[See details](#)

Versions

Ver.1 2023-11-21 01:37:37.765210
[Show All versions](#)

Share

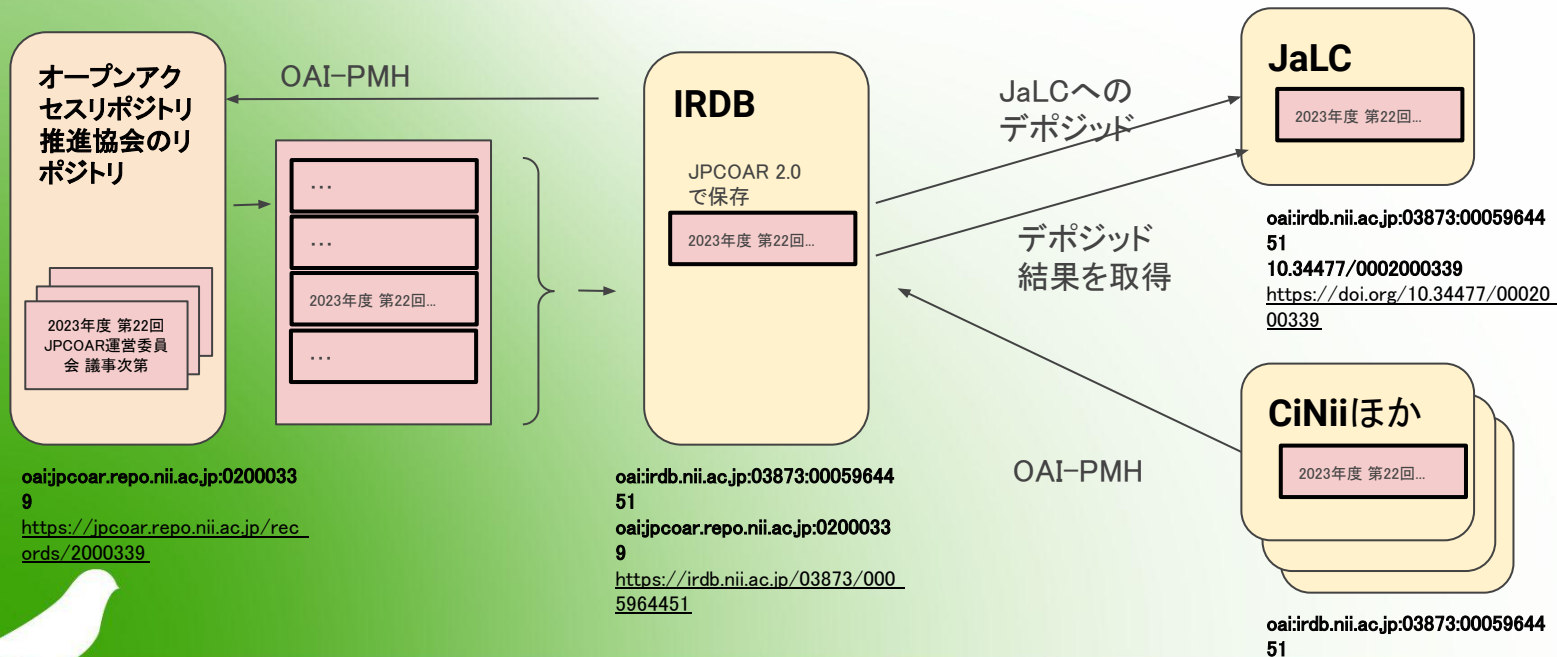


Cite as

n.d., 2023年度 第22回JPCOAR運営委員会 議事次第: オープンアクセスリポジトリ推進協会.

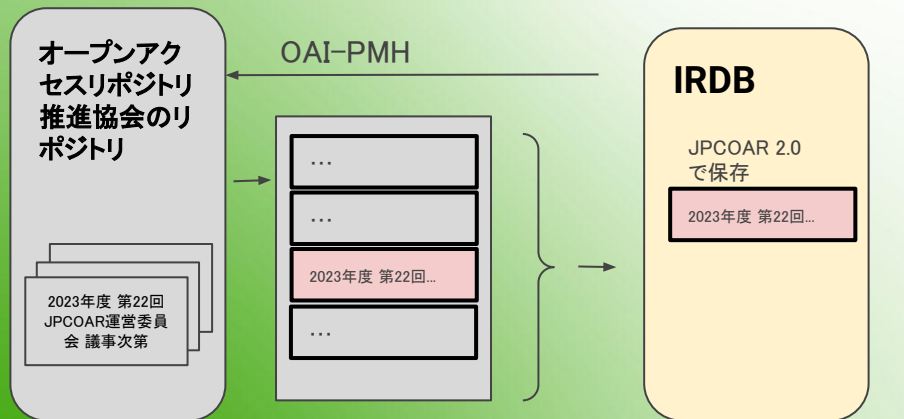
どのように(収集)

「2023年度 第22回JPCOAR運営委員会 議事次第」のアイテムについての流れ



どのように(収集)

「2023年度 第22回JPCOAR運営委員会 議事次第」のアイテムについての流れ



[oai.jpcoar.repo.nii.ac.jp:02000339](https://oai.jpcoar.repo.nii.ac.jp/02000339)
<https://jpcoar.repo.nii.ac.jp/records/2000339>

[oai:irdb.nii.ac.jp:03873:0005964451](https://oai.irdb.nii.ac.jp/03873:0005964451)
[oai.jpcoar.repo.nii.ac.jp:02000339](https://oai.jpcoar.repo.nii.ac.jp/02000339)
<https://irdb.nii.ac.jp/03873/0005964451>

どのように(収集)

OAI-PMHで受信したJPCOAR1.0の形式のXMLを受信

```
<jpcoar:jpcoar>
  <dc:title xml:lang="ja">2023年度 第22回JPCOAR運営委員会 議事次第</dc:title>
  <dc:publisher xml:lang="ja">オープンアクセスリポジトリ推進協会 </dc:publisher>
  <datacite:date dateType="Issued">2023-11-21</datacite:date>
  <dc:language>jpn</dc:language>
  <dc:type rdf:resource="http://purl.org/coar/resource_type/c_1843">other</dc:type>
  <jpcoar:identifier identifierType="DOI">https://doi.org/10.34477/0002000339</jpcoar:identifier>
  <jpcoar:identifier identifierType="URI">https://jpcoar.repo.nii.ac.jp/records/2000339</jpcoar:identifier>
  <jpcoar:identifierRegistration identifierType="JaLC">10.34477/0002000339</jpcoar:identifierRegistration>
  <jpcoar:file>
    <jpcoar:URI objectType="other">https://jpcoar.repo.nii.ac.jp/record/2000339/files/第22回 JPCOAR運営委員会議事次第 .pdf</jpcoar:URI>
    <jpcoar:mimeType>application/pdf</jpcoar:mimeType>
    <jpcoar:extent>247 KB</jpcoar:extent>
    <datacite:date dateType="Available">2023-11-21</datacite:date>
    <datacite:date dateType="Issued">2023-11-21</datacite:date>
  </jpcoar:file>
</jpcoar:jpcoar>
```

参考: https://jpcoar.repo.nii.ac.jp/oai?verb=GetRecord&metadataPrefix=jpcoar_1.0&identifier=oai:jpcoar.repo.nii.ac.jp:02000339

どのように(収集)

プログラムの際にはXMLの階層および属性があるものを1項目毎に処理していくため、複雑な処理となっています。

言語区分のように統一的に扱えるものであれば、複数の項目にまたがって処理できますが、個別の項目に対応するためには、それぞれに処理が必要となります。

どのように(収集)

JPCOAR 1.0で受信したXMLを第1階層別に分解

```
<dc:title xml:lang="ja">2023年度 第22回JPCOAR運営委員会 議事次第</dc:title>
```

```
<dc:publisher xml:lang="ja">オープンアクセスリポジトリ推進協会 </dc:publisher>
```

```
<datacite:date dateType="Issued">2023-11-21</datacite:date>
```

```
<dc:language>jpn</dc:language>
```

```
<dc:type rdf:resource="http://purl.org/coar/resource_type/c_1843">other</dc:type>
```

```
<jpcoar:identifier identifierType="DOI">https://doi.org/10.34477/0002000339</jpcoar:identifier>
```

```
<jpcoar:identifier identifierType="URI">https://jpcoar.repo.nii.ac.jp/records/2000339</jpcoar:identifier>
```

```
<jpcoar:identifierRegistration identifierType="JaLC">10.34477/0002000339</jpcoar:identifierRegistration>
```

```
<jpcoar:file>
```

```
<jpcoar:URI objectType="other">https://jpcoar.repo.nii.ac.jp/record/2000339/files/第22回 JPCOAR運営委員会議事次第 .pdf</jpcoar:URI>
```

```
<jpcoar:mimeType>application/pdf</jpcoar:mimeType>
```

```
<jpcoar:extent>247 KB</jpcoar:extent>
```

```
<datacite:date dateType="Available">2023-11-21</datacite:date>
```

```
<datacite:date dateType="Issued">2023-11-21</datacite:date>
```

```
</jpcoar:file>
```

どのように(収集)

```
<dc:title xml:lang="ja">2023年度 第22回JPCOAR運営委員会 議事次第</dc:title>
```

■ チェック処理

- 1.スキーマ全体で dc:title が1つ以上あるか？(必須チェック)
- 2.xml:langが正しいか？
不正な言語区分は、言語区分をカットして登録
- 3.同じ言語区分で dc:title が複数ないか？

どのように(収集)

```
<dc:publisher xml:lang="ja">オープンアクセスリポジトリ推進協会 </dc:publisher>
```

■ チェック処理

1.xml:langが正しいか？

- ・不正な言語区分は、言語区分をカットして登録

どのように(収集)

```
<datacite:date dateType="Issued">2023-11-21</datacite:date>
```

■ チェック

- 1.dateTypeの値が正しいか？
- 2.日付の形式が正しいか？
 - ・/(スラッシュ) や .(ピリオド) は、-(ハイフン)に変換
 - ・年月日、年月、年のみ形式の妥当性チェック
 - ・月や日が00であれば、削除
 - ・時刻があると時刻をカット
 - ・年が和暦の場合は、西暦に変換

どのように(収集)

```
<jpcoar:identifier identifierType="DOI">https://doi.org/10.34477/0002000339</jpcoar:identifier>  
<jpcoar:identifier identifierType="URI">https://jpcoar.repo.nii.ac.jp/records/2000339</jpcoar:identifier>
```

```
<jpcoar:identifierRegistration identifierType="JaLC">10.34477/0002000339</jpcoar:identifierRegistration>
```

■ チェック

- 1.identifierのDOIとidentifierRegistrationのDOI部分が一致しているかどうか？
- 2.DOI以外(URIまたはHNL)のidentifierが登録されているかどうか？
- 3.機関情報のDOIが、10.34477 と一致しているかどうか？
- 4.10.34477/0002000339のDOIが以前に登録されているかどうか？
登録されている場合は、IDが一致しているかどうか

* DOIは、一度有効化された場合は、別のアイテムに付け替えができない。

どのように(収集)

```
<jpcoar:file>  
<jpcoar:URI objectType="other">https://jpcoar.repo.nii.ac.jp/record/2000339/files/第22回 JPCOAR運営委員会議事次第 .pdf</jpcoar:URI>  
<jpcoar:mimeType>application/pdf</jpcoar:mimeType>  
<jpcoar:extent>247 KB</jpcoar:extent>  
<datacite:date dateType="Available">2023-11-21</datacite:date>  
<datacite:date dateType="Issued">2023-11-21</datacite:date>  
</jpcoar:file>
```

■ チェック

- 1.URIの妥当性チェック
- 2.mimeTypeの存在性チェック
- 3.dateTypeの属性の存在チェック
- 4.dateの日付の内容チェック

どのように(収集)

すべての項目についてチェックが完了して、始めてIRDBにデータが登録されます。

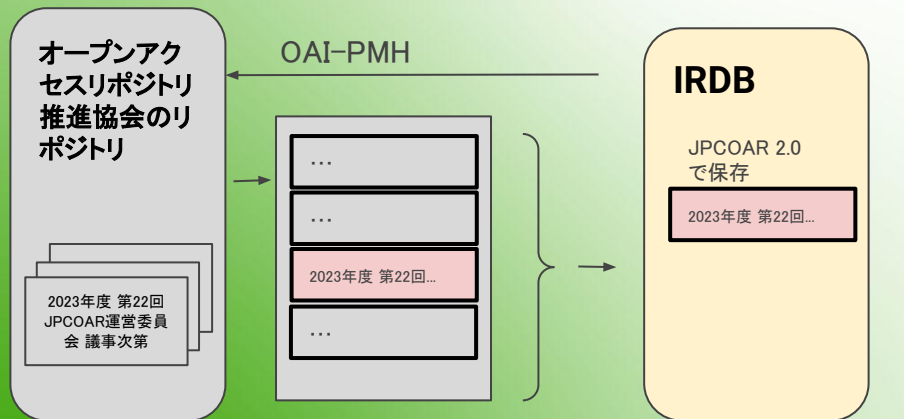
登録する場合には、IRDB側では全体のデータを管理するための管理IDが付与されて、アイテムに対してIDで管理されます。

IRDBのID : 03873:0005964451

リポジトリ側のID : oai:jpcoar.repo.nii.ac.jp:02000339

どのように(収集)

「2023年度 第22回JPCOAR運営委員会 議事次第」のアイテムについての流れ



oai.jpcoar.repo.nii.ac.jp:02000339
<https://jpcoar.repo.nii.ac.jp/records/2000339>

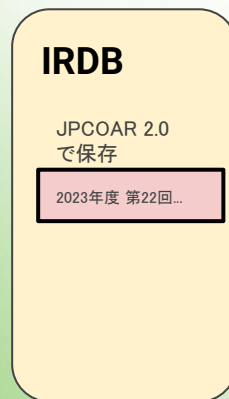
oai.irdb.nii.ac.jp:03873:0005964451
oai.jpcoar.repo.nii.ac.jp:02000339
<https://irdb.nii.ac.jp/03873/0005964451>

どのように(提供)

- ② 収集したメタデータのチェックを行い外部に提供
(CiNii、国立国会図書館、ディスカバリーサービスなど)

どのように(提供)

「2023年度 第22回JPCOAR運営委員会 議事次第」のアイテムについての流れ



`oai:irdb.nii.ac.jp:03873:00059644`
`51`
`oai:jpcoar.repo.nii.ac.jp:0200033`
`9`
<https://irdb.nii.ac.jp/03873/0005964451>

OAI-PMH



`oai:irdb.nii.ac.jp:03873:00059644`
`51`

どのように(提供)

IRDBで登録されたデータは、外部(CiNii、国立国会図書館、ディスカバリーサービス)のサービスにデータ提供を実施しています。

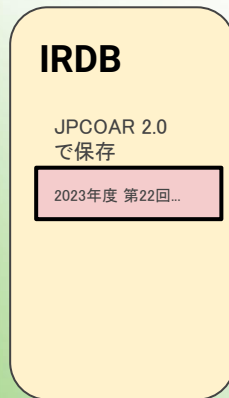
各サービスは、OAI-PMHのプロトコルでデータ収集されています。

データ収集は、各サービスから任意のタイミングでデータが取得されています。取得周期は各サービスによって異なります。

また、データは、新規登録・更新があったデータのみを取得しています。さらに、国立国会図書館であれば博士論文のデータのみ取得するなど部分的なデータ提供も行っています。

どのように(収集)

「2023年度 第22回JPCOAR運営委員会 議事次第」のアイテムについての流れ



[oai:irdb.nii.ac.jp:03873:0005964451](https://irdb.nii.ac.jp/03873/0005964451)
51
[oai:jpcoar.repo.nii.ac.jp:02000339](https://irdb.nii.ac.jp/03873/0005964451)
9
<https://irdb.nii.ac.jp/03873/0005964451>

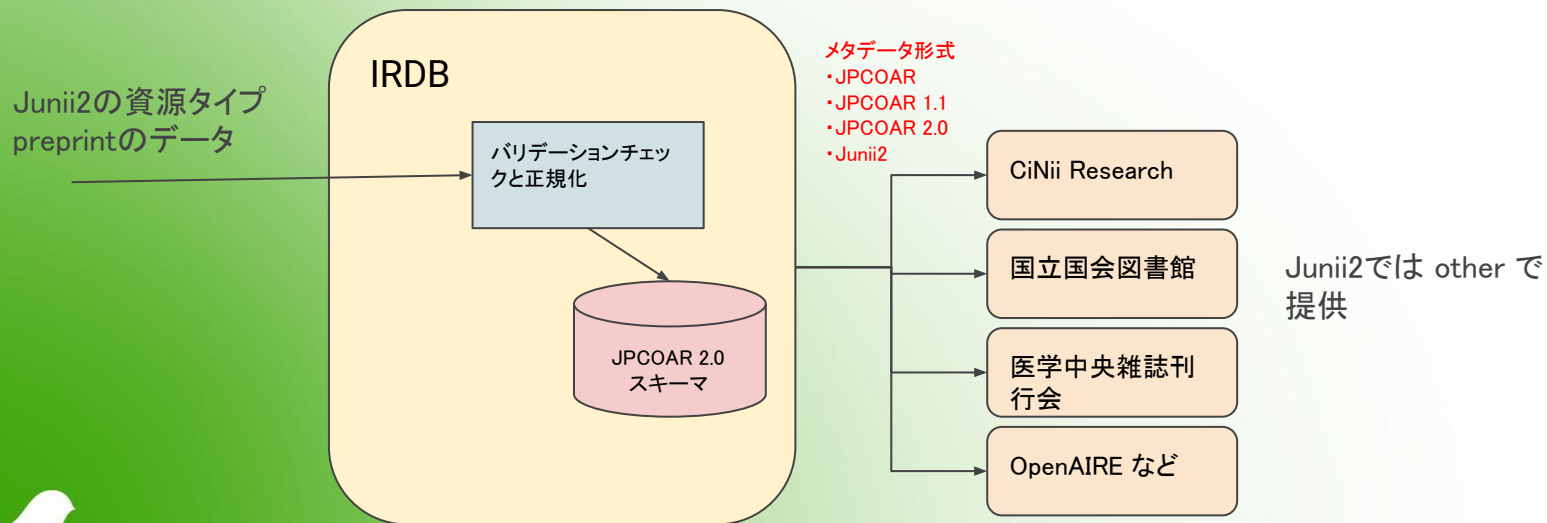
OAI-PMH



[oai:irdb.nii.ac.jp:03873:0005964451](https://irdb.nii.ac.jp/03873/0005964451)
51

どのように(提供)

外部へのデータ提供スキーマについては、junii2、jpcoar、jpcoar_1.1、jpcoar_2.0形式で提供しています。

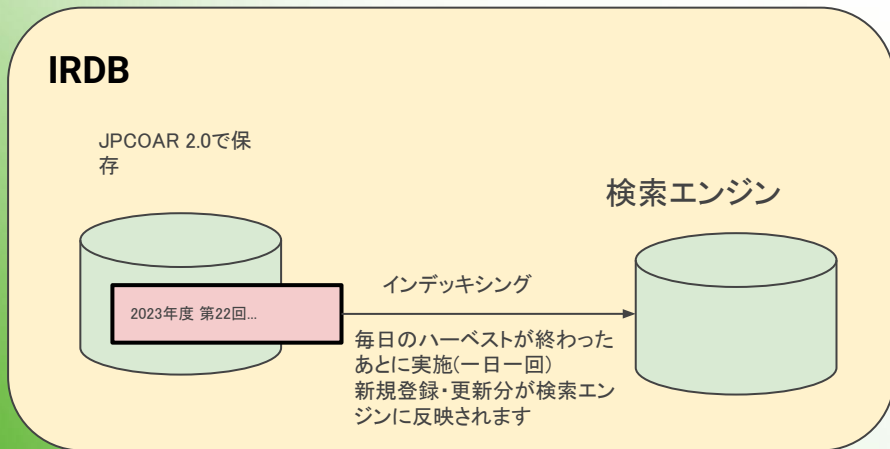


どのように(検索)

- ① 収集したメタデータを一括で検索

どのように(収集)

「2023年度 第22回JPCOAR運営委員会 議事次第」のアイテムについての流れ



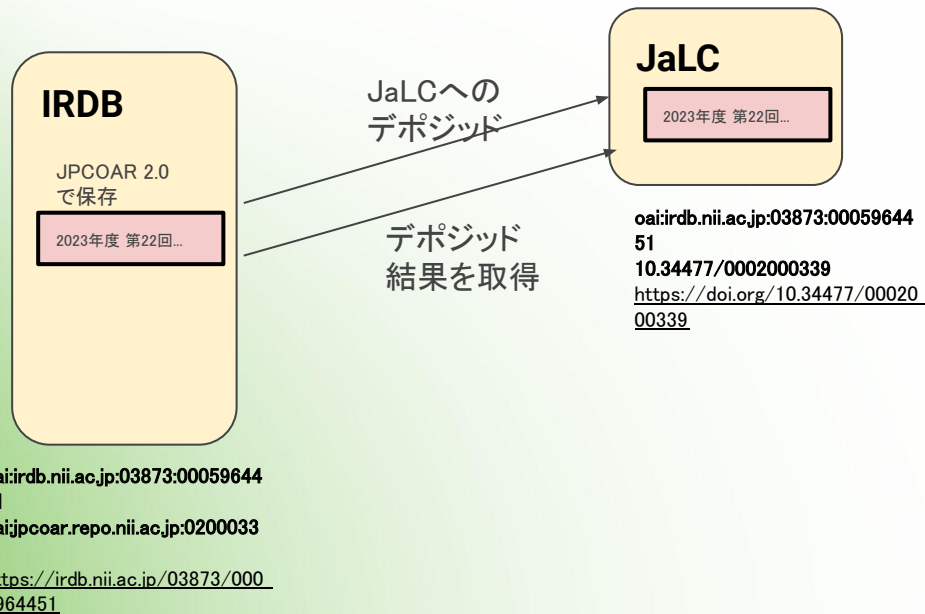
oai:irdb.nii.ac.jp:03873:0005964451
<https://irdb.nii.ac.jp/03873/0005964451>

どのように(DOI有効化)

③ DOIの有効化(JaLCとの連携)

どのように(DOI有効化)

「2023年度 第22回JPCOAR運営委員会 議事次第」のアイテムについての流れ



どのように(DOI有効化)

IRDBに登録されDOIがあるものについて、毎日JaLCにWebAPIを利用して、DOIの有効化処理がされます。

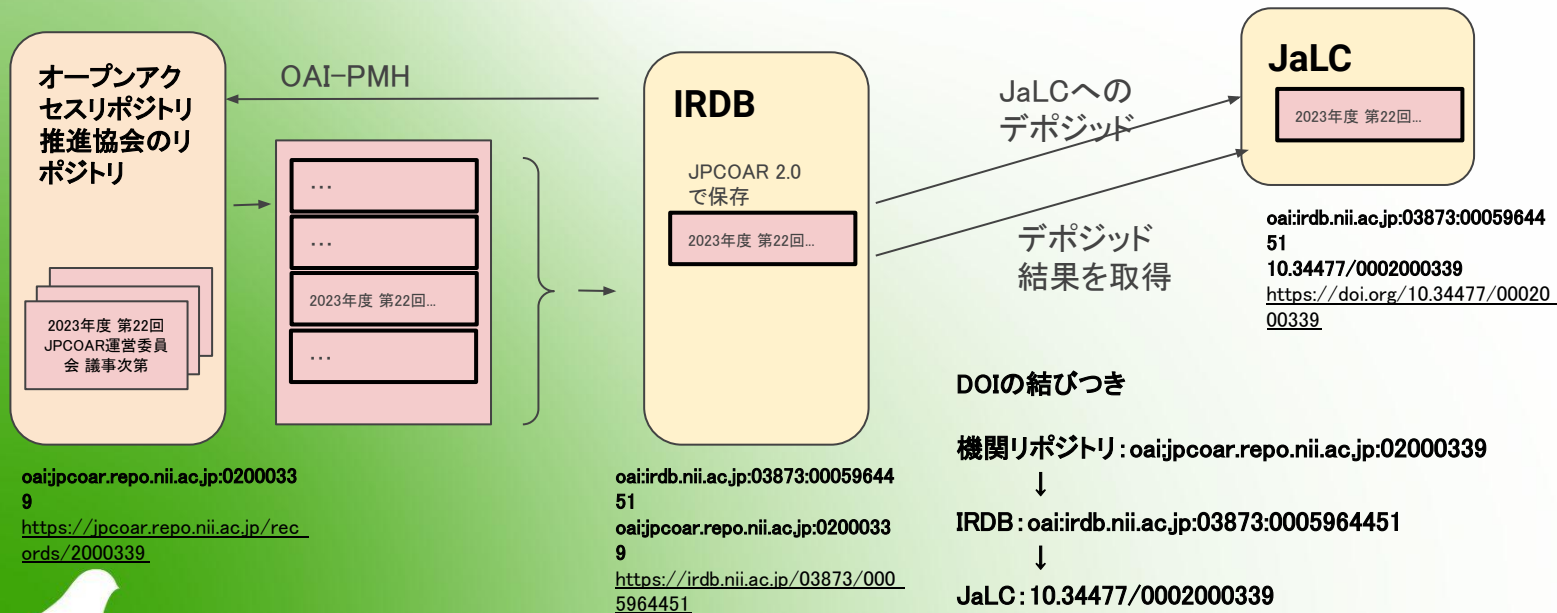
JaLC側でDOIの有効化処理されることで、各機関リポジトリで付与されたDOIがDOIとして利用が可能となります。

一度付与されたDOIは変更できず、登録したアイテムとDOIは永続的に結びつきされます。

* 同じアイテムの認識は機関リポジトリ側でのIDによる判断

どのように(DOI有効化)

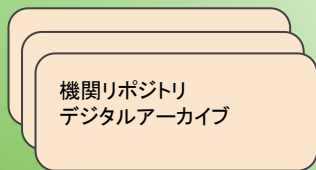
「2023年度 第22回JPCOAR運営委員会 議事次第」のアイテムについての流れ



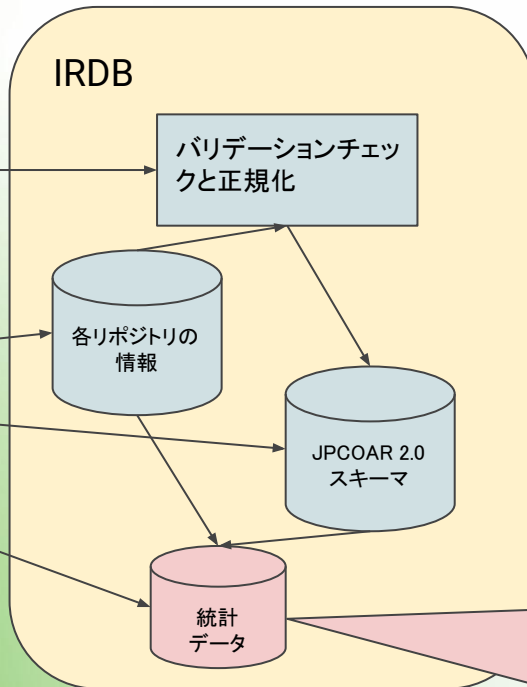
どのように(統計情報)

④ 月次での統計情報の作成(全体、リポジットリ毎)

どのように(統計情報)



ご担当者様



毎月1日
リポジトリ単位にIRDBに登録されているデータから資源タイプ毎に集計を行っています。
その時点の統計データということで内部で保持しております。

まとめ

まとめ

「IRDBのカラクリと役割 ～ どこから・どこへ・どのように」についてご説明させていただきました。

IRDBは国内の各リポジトリにあるメタデータを集約して、世界に流通させる中心的な役割を担っており、またDOIの有効化の処理という重要な機能があります。華やかな注目されるシステムではないですが、日々必要なシステムと考えています。

サン＝テグジュペリの「星の王子さま」にある「本当に大切は物は目に見えない」のように、IRDBはその言葉を表すシステムかと思っています。

そして、そのデータの発生源として、日々皆様がデータを登録されていることで成り立っています。今日、この日からIRDBのことを理解いただき、皆さんの業務が大切なことだと再認識いただければ幸いです。

ご清聴ありがとうございました