

# ハーベストの概略とIRDB の基本的な活用

第4回学術コミュニケーションセミナー  
IRDB-カラクリと役割:どこから・どこへ・どのように-

信州大学附属図書館 伊東洋輔

2024.01.17



# リポジトリにおけるハーベスト（収穫）とは

機関リポジトリに登録したメタデータやアイテムを異なるデータベースでも利用できるようにする・リンクさせる行為。

- ハーベスト＝「（農産物を）収穫する」  
機関リポジトリの実りを外部サイトが刈り取っていくイメージ。  
と、言うとは悪印象だが、現実と違って収穫されてもモノは無くなる。アクセスルートが純増するので、基本的に悪い事ではない。
- スキーマとは  
メタデータ入力規則。誤解を恐れずに言うと、リポジトリ等のデータベースにおいてシソーラスのような役割を果たしている。

※参考

IRDB ハーベスト仕様

<https://support.irdb.nii.ac.jp/ja/harvest>

JPCOARスキーマガイドライン

<https://schema.irdb.nii.ac.jp/ja>

# ハーベストの概念図

## 各機関

コンテンツを登録し、  
閲覧に供すると同時に  
ハーベストによる  
データ提供をおこなう



## ハーベスト

### 共同リポジトリ



### IRDB



IRDBは公開基盤というよりは  
データベース間の仲介役

## 閲覧者へ



### CiNii、NDLなど



# ハーベスト処理結果の通知メール

件名：ハーベスト処理結果の通知メールです。

ハーベスト処理結果の通知メールです。

信州大学機関リポジトリ BaseURL : <https://xxxxxxx.repo.nii.ac.jp/oai>

前回ハーベスト開始日時：2023-11-28 18:07:49

前回ハーベスト終了日時：2023-11-28 18:07:52

ハーベスト対象データ件数：3 登録件数：2 更新件数：1 削除件数：0 データ取得エラー：0

レコードエラー件数：0 項目エラー件数：3 ワーニング件数：6 項目変換件数：0

確認先URL : <https://irdb.nii.ac.jp/usercontents>

■=■+■：前回から変化があったアイテム数。  
■：アイテム自体のハーベスト回避になるエラーがあったアイテム。メールにエラー内容が記載される。  
■：IRDBのサイトで内容を確認する。エラー項目はハーベストされないので空値or前回ハーベスト時のまま、ワーニングは不備はあるもののハーベストされている。

■・■・■はアイテム数、■は項目数なので  
■<■となることもある。

ここからハーベスト結果の詳細を確認できる。  
(要ID・パスワード)

## ●エラー/ワーニングになったアイテムの再ハーベストについて

ハーベスト対象になるアイテムは、新規登録/最終更新日時が  
前回のハーベスト日時より新しいもの。

エラー/ワーニングになったアイテムでも、  
修正が行われるまでは次回以降ハーベスト対象にならない。

ハーベスト対象データ件数：3 登録件数：2 更新件数：1 削除件数：0 データ取得エラー：0  
レコードエラー件数：0 項目エラー件数：3 ワーニング件数：6 項目変換件数：0

エラーが有っても修正しなければ翌週は



ハーベスト対象データ件数：0 登録件数：0 更新件数：0 削除件数：0 データ取得エラー：0  
レコードエラー件数：0 項目エラー件数：0 ワーニング件数：0 項目変換件数：0

# ハーベスト結果確認画面

oai-id	OAI連携先種別	DOI
<input type="text"/>	<input type="text" value="- すべて -"/>	<input type="text" value="- すべて -"/>
更新日付(開始日)	更新日付(終了日)	メッセージ
<input type="text"/>	<input type="text"/>	<input type="text" value="- すべて -"/>
項目エラー件数(以上を検索)	ワーニング件数(以上を検索)	項目変換件数(以上を検索)
<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="button" value="🔍 検索"/>		

●更新日付（開始日） & （終了日）  
対象日（期間）を挟むように指定する。（YYYYMMDD）

ex.2023年12月6日がハーベスト日のアイテムを見たい場合  
『開始日：20231205 終了日：20231207』

●エラー/ワーニング件数  
空欄なら正常も含むハーベスト結果が、どちらかに（両方に）「1以上の数字」で検索するとエラー/ワーニングがあったアイテムだけを表示できる。

# IRDB : 統計情報の確認機能

エラー/ワーニングの確認は「マイコンテンツ」

対象年

2023

対象月

11

検索

特定の年月におけるハーベスト状況を確認できる。

統計ファイル (全機関の合計値)

統計ファイル (全機関の機関別統計)

総件数	本文あり件数(率)	DOI付与件数(率)
21,523	21,517 (99.97 %)	523 (2.43 %)

すべてのデータ

資源タイプ別コンテンツ数

conference paper

132

departmental bulletin paper

8,263

accessRights別コンテンツ数

metadata only access

open access

オブジェクトタイプ別コンテンツ数

fulltext

summary

7

# IRDB : 統計情報の確認機能

## コンテンツ統計(ユーザ)中段

### すべてのデータ

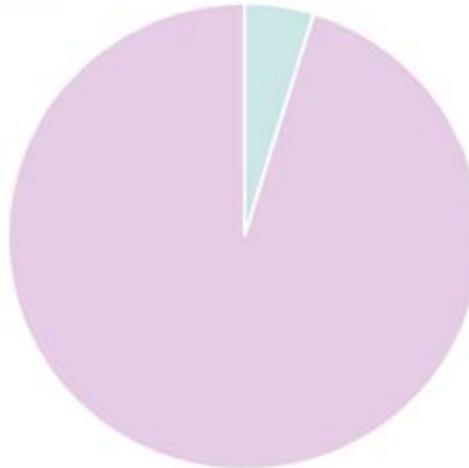
#### 資源タイプ別コンテンツ数

conference paper	132
departmental bulletin paper	8,264
journal article	7,998
article	336
book	123
conference output	84
dataset	10
research report	2,841
technical report	18
thesis	2
doctoral thesis	1,090
learning object	6
other	792

ハーベスト済アイテムの分類別数一覧。(≠リポジトリアイテム総数)  
特定の時期の数も出せるので、「ある月-1年前」で年間純増数も算出可

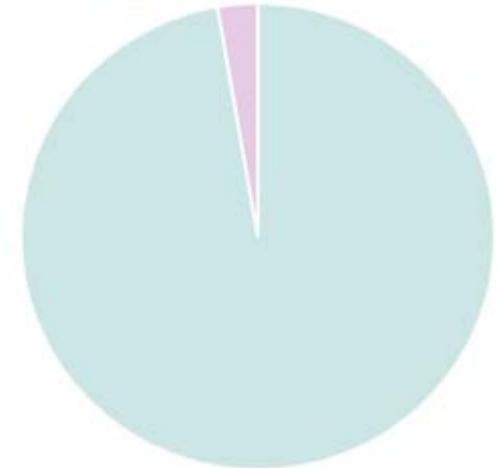
#### accessRights別コンテンツ数

metadata only access    open access



#### オブジェクトタイプ別コンテンツ数

fulltext    summary



アクセス権、アイテム情報のオブジェクトタイプ別の分類  
エンバーゴ付きで公開したアイテムの更新忘れなどの確認に

# IRDB : 統計情報の確認機能

## コンテンツ統計 (ユーザ) 下段

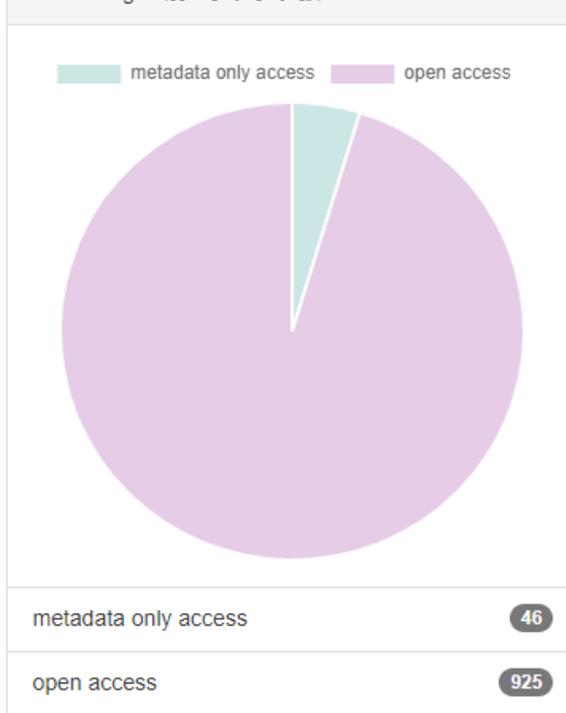
本文ありのデータ

本文未登録・未公開を除外した数も確認できる。  
メタデータのみ登録したアイテムの確認などに。

### 資源タイプ別コンテンツ数

conference paper	132
departmental bulletin paper	8,264
journal article	7,993
article	336
book	123
conference output	84
dataset	10
research report	2,840
technical report	18
thesis	2
doctoral thesis	1,090
learning object	6
other	792

### accessRights別コンテンツ数



### オブジェクトタイプ別コンテンツ数

