

## 情報サービス：その裏側の技術から

国立研究開発法人 国際農林水産業研究センター（国際農研）  
情報広報室 広報資料科 情報高度利用専門職

林賢紀

tzhaya@affrc.go.jp



2021年11月5日(金) 私立大学図書館協会2021年度研修分科会

1

1

## 国際農研について

熱帯及び亜熱帯に属する地域その他開発途上地域における農林水産業に関する技術向上のための試験研究を行う国立研究開発法人

- 研究者の派遣・招へいによる国際共同研究の推進
- 海外情報の収集・分析・提供と広報
- 国際シンポジウム・ワークショップ等の開催
- 政府や他機関との科学技術に関する国際連携や交流
- つくば市、石垣市に所在
- 研究職員139名、一般職員等39名



2

## 自己紹介

- 略歴
  - 1993-2013 農林水産省（農林水産研究情報総合センター）
  - 2014-(現職) 国立研究開発法人国際農林水産業研究センター
  - 2016 博士（情報学）
- 所属
  - 情報広報室 広報資料科
    - 広報、刊行物、Webサイト、図書館などの担当部署
- 業務
  - 図書館の管理運営
    - 分類でいえば専門図書館
    - 職員は林と契約職員1名
  - 公式Webサイトの管理運営
    - システムのメンテナンス
    - その他ページの更新作業など
- 委員など
  - これからの学術情報システム構築検討委員会
  - 図書館システム・ネットワーク：ワークフロー検討作業部会委員



3

## 本日のテーマと講義内容

これまで電子書籍、機関リポジトリ等で大学図書館が情報を提供・管理する側面をテーマに業務を見つめてきました。第5回では利用者の立場から情報サービスについて考えたいと思います。図書館業務のスキルは、「読み・書き」の両面からアプローチが必要です。情報管理に必要なメタデータ作成の「書く」技術の裏に、それが利用者にどう使われるか想像できることがベースにあり、一方レファレンスやILLで的確に情報サービスができるようになる「読む」技術を鍛えるには、**データがどこから来て、どのように作成されている**か知っておくことが役に立ちます。GoogleとWikipediaで欲しい情報が簡単に手に入る世の中で、質問に来た利用者に満足いく回答ができるようになるには、常に最新の情報サービスのツールを使いこなせるようになるのが理想ですね。また、利用者が求める資料（資源）を入手するまでが情報サービスですので、レファレンスとILLやレフェラルサービスについて、考える機会にしたいと思います。

(開催案内より)



4

## 情報サービスの裏側を考える

- そのデータはどこから来ているの？
  - いろいろな情報をまとめて検索できる
- どうやって連携しているの？
  - 検索結果から電子ジャーナルの全文にリンクがある
- どんな技術が使われている？



5

## 例えば医中誌の紹介から

これらの機能はどうやって実現されている……？

### オンラインジャーナルへのリンク

- 「医中誌Web」の検索結果には、各種オンラインジャーナルやPubMedなどへのリンクアイコンが表示されます。リンク先のサービスを契約している場合、あるいは論文が無料で公開されている場合には、アイコンをクリックするとその場ですぐに論文を読むことができます。

### 図書館システムへのリンク

- 法人様にご提供する「医中誌Web」では、ご利用の機関にて設定を行うと、検索結果にOPACやリンクリゾルバへのリンクアイコンが表示され、検索に続きすぐに所蔵の確認が行えます。

医中誌Webとは  
<https://www.iamas.or.jp/service/ichu/>より

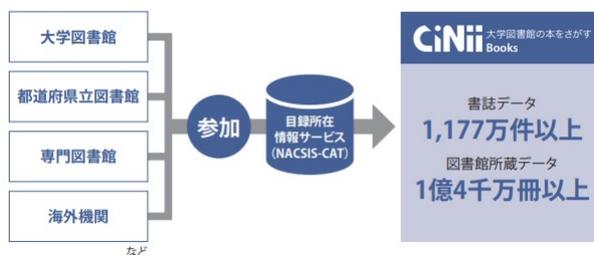


6

## 情報サービスの例

## CiNii Books

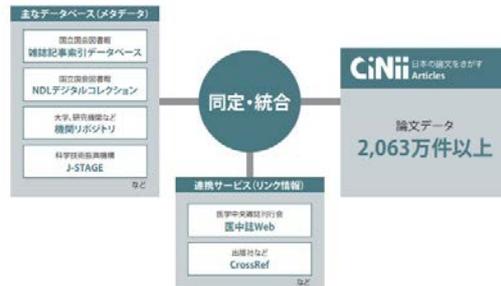
- 全国の大学図書館等が所蔵する本（図書・雑誌）の情報を検索
- 目録所在情報サービス(NACSIS-CAT)のデータを元にしている
- 国内の電子ブック、電子ジャーナルなど電子リソースにもリンク



CiNiiパンフレット（2018年4月）より

## CiNii Articles

- 学協会刊行物・大学研究紀要・国立国会図書館の雑誌記事索引データベースなどの学術論文情報を検索
- 全文情報やCiNii Booksともリンク



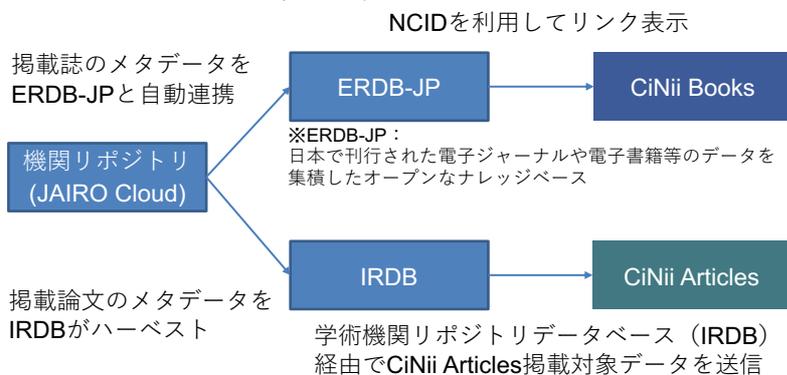
CiNiiパンフレット (2018年4月) より



9

## CiNii Books/Articleと電子リソースのリンク (例)

- 機関リポジトリ (JAIRO Cloud)、ERDB-JPとリンク



10

## JAIRO Cloud-ERDB-JP-CiNii連携

- 書誌レベルの情報をERDB-JPが取得、CiNii Booksと連携できるデータにしている

### JAIRO Cloud x ERDB-JP

JAIRO Cloudに登録した雑誌情報は、自動的にERDB-JPにも反映され、OPAC、リンクリゾルバ、ディスカバリサービスなど、世界中のサービスから検索されます。



•ERDB-JPリーフレット 2018 JAIRO Cloud自動連携.pdf より



11

## CiNii Research

- CiNii Articles/Books/Dissertationsと研究データ、KAKENをまとめて検索
  - 研究データはIRDB（さらに元は機関リポジリ上のデータ）が対象
- 2022年4月1日より、CiNii Articlesは、CiNii Researchへ統合
  - 論文検索はCiNii Researchに一本化



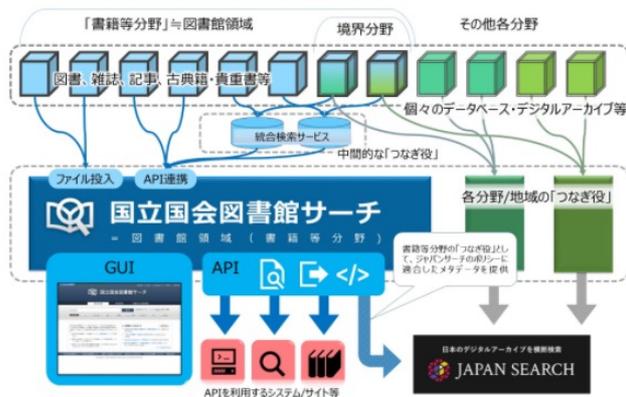
12

## 国立国会図書館サーチ

- 国立国会図書館や他の機関が保有する、蔵書目録、デジタル資料、立法情報、レファレンス事例や出版情報など、多様な情報を統合的に検索
- 総合目録ネットワーク（ゆにかねっと）（公共図書館蔵書）、農林水産関係試験研究機関総合目録、CiNii など多数の連携先がある

## 国立国会図書館サーチの連携イメージ

- ファイル投入、API連携等により統合的な検索を行う

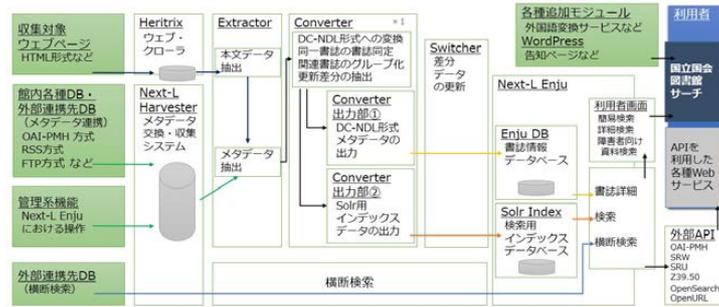


国立国会図書館サーチ連携拡張に係る実施計画（2021-2025版）  
（別紙）国立国会図書館サーチの連携イメージ

## 国立国会図書館サーチのデータの流れ

- 様々な方法でデータを取得、または横断検索

国立国会図書館サーチのデータの流れ

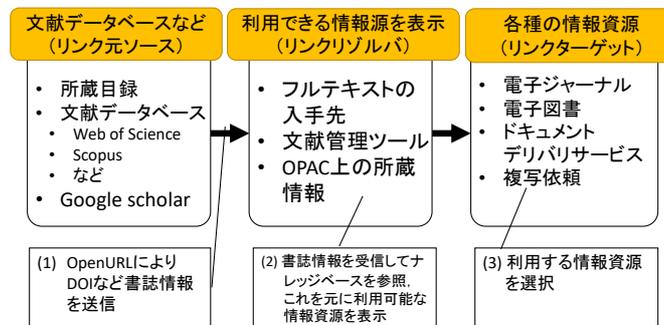


国立国会図書館サーチ システムについて  
<https://iss.ndl.go.jp/information/outline/system/>



## リンクリゾルバ

- 文献データベースと利用できる情報資源を結びつける
  - 図書館では電子ジャーナルの購読状況やURLなど、リンクする情報資源を登録する
- SFX、360Linkなどの製品がある
- 「ナレッジベース」を参照して電子ジャーナルのURLを検索
  - このナレッジベースにERDB-JPなどのデータを利用できる



## ディスカバリサービス

- Primo、Summon (ExLibris) など
- 「Central Index」と呼ばれる国内、国外の出版社やアグリゲーターから収集された学術情報を検索対象とする
  - どの出版社やアグリゲーターを検索対象とできるかがポイント
  - 各ディスカバリサービス用にメタデータをそろえて収録されている
- 自館OPACの書誌・所蔵情報や電子ジャーナルの購読情報も検索・参照できる



17

## Google

- 全世界のWebページを収集
- ページ内のメタデータを元にインデックス
  - 「イベント」「レシピ」「Q&A」など記載内容により検索結果の見せ方を変える
  - HTMLタグなどで記述された情報を処理
  - 詳細は「構造化データの仕組みについて」を <https://developers.google.com/search/docs/advanced/structured-data/intro-structured-data?hl=ja>

ページの意図を伝える明示的な手がかりとして構造化データを提供してもらくと、Googleはそのページをより正確に理解できるようになります。構造化データとは、ページに関する情報を提供し、そのコンテンツ（たとえば、レシピページの場合は材料、加熱時間と加熱温度、カロリーなど）を分類するために標準化されたデータ形式です。

18

## イベントが記述されたHTMLページの例

- この例はJSON-LDで記述

<pre>&lt;html&gt; &lt;head&gt; &lt;title&gt;The Adventures of Kira and Morrison&lt;/title&gt; &lt;script type="application/ld+json"&gt; {   "@context": "https://schema.org",   "@type": "Event",   "name": "The Adventures of Kira and Morrison",   "startDate": "2023-07-21T19:00-05:00",   "endDate": "2025-07-21T23:00-05:00",   "eventAttendanceMode": "https://schema.org/OfflineEventAttendanceMode",   "eventStatus": "https://schema.org/EventScheduled",   "location": {     "@type": "Place",     "name": "Snickerpark Stadium",     "address": {       "@type": "PostalAddress",       "streetAddress": "100 West Snickerpark Dr",       "addressLocality": "Snickertown",       "postalCode": "19019",       "addressRegion": "PA",       "addressCountry": "US"     }   } }, </pre>	このページは イベント情報である
	イベント名
	開始日時
	場所
	住所



19

## Google Scholar

- 学術論文のみを検索できる
- 収集対象がGoogleとは異なる
  - 特定のWebサイトのページ（一定の基準がある）
  - 特定のHTMLタグからインデックス（指定の記述方法がある）
- リンクリゾルバのデータと連携
  - リンクリゾルバ上の電子ジャーナル購読情報
  - 所属組織のIPアドレス
- CASA: Campus Activated Subscriber Access
  - 購読している電子ジャーナルへ「Google Scholar経由で」リモートアクセス

20

## Google Scholarに対応したページの例

- 各出版社、J-STAGEなど、各論文のページにこのような記述を含んでいる

```
1 <?xml version="1.0" encoding="UTF-8" ?>
2 <html lang="ja" xmlns:xm="http://www.w3.org/1998/Math/MathML">
3 <head>
4
5
6
7 <!-- 2.0.15 -->
8 <meta charset="utf-8" />
9 <meta http-equiv="X-UA-Compatible" content="IE=edge" />
10 <meta name="format-detection" content="telephone=no" />
11 <meta name="viewport" content="width=device-width, initial-scale=1.0, maximum-scale=1.0, user-scalable=no" />
12 <meta name="description" content="J-STAGE" />
13 <meta name="citation_journal_title" content="情報知学会誌" />
14 <meta name="citation_journal_abbrev" content="情報知学会誌" />
15 <meta name="citation_journal_abbrev" content="Journal of Japan Society of Information and Knowledge" />
16 <meta name="citation_publisher" content="情報知学会" />
17 <meta name="citation_author" content="林 真樹" />
18 <meta name="citation_author_institution" content="国際農林水産業研究センター" />
19 <meta name="citation_author" content="海保 隆一郎" />
20 <meta name="citation_author" content="長岡 晋男" />
21 <meta name="citation_author_institution" content="筑波大学 図書館情報メディア系 / 知的コミュニティ基礎研究センター" />
22 <meta name="citation_title" content="複合的な情報資源のLinked Open Dataにおける構造分析とスキーマ定義手法" />
23 <meta name="citation_publication_date" content="2016/02/25" />
24 <meta name="citation_online_date" content="2016/04/15" />
25 <meta name="citation_volume" content="26" />
26 <meta name="citation_issue" content="1" />
27 <meta name="citation_firstpage" content="11" />
28 <meta name="citation_lastpage" content="26" />
29 <meta name="citation_doi" content="10.2864/jisik.2016.001" />
30 <meta name="citation_fulltext_world_readable" content="" />
31 <meta name="citation_pdf_url" content="https://www.jstage.jst.go.jp/article/jisik/26/1/26_2016_001/pdf" />
32 <meta name="citation_issn" content="0917-1436" />
33 <meta name="citation_isbn" content="1881-7881" />
34 <meta name="citation_language" content="ja" />
35 <meta name="citation_from_crossref" content="1,linked Open Data" />
```



21

## 参考：Google Scholar Campus Activated Subscriber Access

- シームレスなリンクおよび認証を提供することで、エンドユーザーの所属機関が契約している学術コンテンツへのリモートアクセスを可能にする機能
  - 多くの出版社が対応
- 利用方法
  - 組織のネットワーク内でGoogle Scholarにアクセス
    - Googleアカウントでログインするか、同じPCを使用する
    - 所属機関の契約と紐づけられてキャンパス外からもアクセスが可能に
    - 機関契約の利用（IPアドレス認証）として識別



22

## 情報サービスを実現するための技術

## メタデータ、スキーマ

- メタデータ？
  - 情報資源の属性値  
たとえば、本に対する目録データ
  - データに関するデータ
  - 「ペットボトルのラベル」
- メタデータスキーマ
  - どの項目に、どんな記述をすればよいか？
  - どのようなメタデータを作成するか、構造的定義を示したもの
  - メタデータを構成するデータ項目とその扱いを定義
  - 検索項目として設定されていなければ検索できない

## メタデータスキーマの例

- JPCOARスキーマ
  - オープンアクセスリポジトリ推進協会（JPCOAR）が策定
  - （JAIR Cloudなど）機関リポジトリの収録対象を記述するためのメタデータスキーマ
  - 日本の機関リポジトリのメタデータの国際的な相互運用性を向上させ、日本の学術的成果の円滑な流通を図ることを目的
  - 論文だけでなく研究データ、デジタルアーカイブ等にも適用できる
- 国立国会図書館ダブリンコアメタデータ記述（DC-NDL）
  - メタデータ記述に用いる語彙の国際標準であるDublin Core（ダブリン・コア）を基に、国立国会図書館独自の拡張を加えたもの
  - 主に国立国会図書館サーチで使用



25

## 検索サービスでの対応

- データベースごとにスキーマ（≒規格）が異なる
  - それぞれのニーズなどによって、検索できる（収録されている）項目は異なる
- 統合するためにはどこかで一つにそろえる必要がある
  - ディスカバリーサービスのCentral Indexもこれを大規模に行っている
  - 主要な項目は一つに揃えやすいが、漏れる項目があるかもしれない

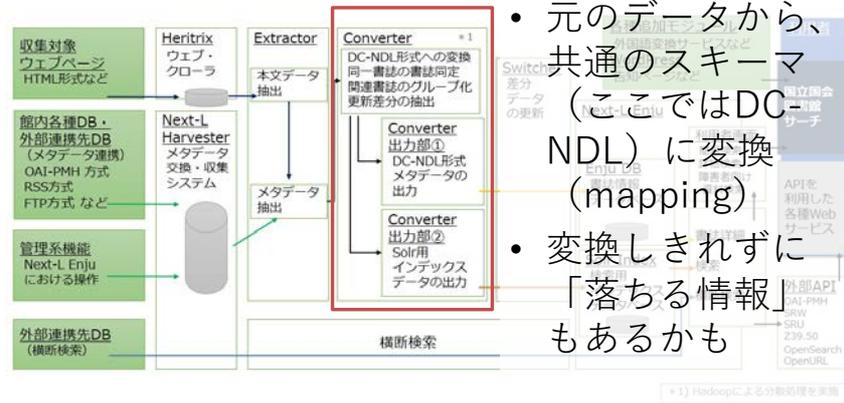
	タイトルの収録項目名	著者名の収録項目名
データベースA	Title	Author
データベースB	TR	AL
データベースC	タイトル	責任表示
データベースD	TL	AU
要検討事項	（翻訳など別タイトルの有無？）	（役割が違うと別項目？）



26

# 国立国会図書館サーチでの変換例

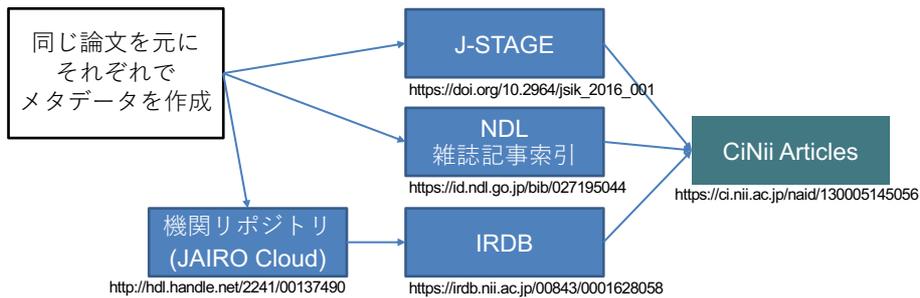
## 国立国会図書館サーチのデータの流れ



- 元のデータから、共通のスキーマ（ここではDC-NDL）に変換（mapping）
- 変換しきれずに「落ちる情報」もあるかも



## 「落とされている」などの情報の例



- CiNii Articlesで3つのデータベースのメタデータを統合する
- 以下はそれぞれ元のデータベースでないと表示されない
  - J-STAGEにあった引用文献
  - NDL雑誌記事索引にあった「件名」
- 相互に補う情報もある
  - NDL雑誌記事索引には著者所属と英語名を採録していないが、他のDBにはあるのでCiNii Articlesでは表示できる



## 参考：文献への索引情報の付与

- 文献からどのような情報を取り出して記録するか？
  - タイトル、著者（著者所属）、掲載誌…
  - 引用文献や資金提供元（多くは謝辞にある）
    - 特に引用文献を重視したのがWeb of Science
  - キーワードを独自に付与する場合も
    - 医中誌は独自のシソーラスを作成、  
人手によるキーワード付け（インデクシング）
- データベースごとに特徴がある
  - 特に有償のデータベースはインデクシングにコストをかけている

## 参考：横断検索とメタデータの収集

- どちらも複数のデータベースを検索するための手段
  - 国立国会図書館サーチは対象によって使い分けている
- 横断検索
  - あるデータベースから、他のデータベースを対象に検索を実行して結果を統合して表示
  - 結果が返ってくるまで待機する必要がある
  - 常に更新された情報が得られる
- メタデータの収集
  - あるデータベースに、他のデータベースのメタデータを取り込んでデータベースを構築する
  - 横断検索より高速
  - データ収集の頻度で更新頻度がかわり、最新のものではない可能性がある

## データを統合したときの問題

- 「名寄せ」の問題
- 「同じ」データは一つにしたい
  - 同じ（ように見える）タイトルのレコードが複数ある
  - 同じ（ように見える）著者名が複数ある
    - 所属が違うが実は同一人物
- CiNiiでは、機械的に名寄せできなかったデータは利用者が報告できる
  - 作成元が異なるデータを統合した際に、判断できない場合がある



31

## どうやってメタデータを 集める・交換するか

- Excelファイルで
  - CSV形式でやりとりする場合も多い
- API (Application Program Interface) で効率的に
  - システム同士でデータをやりとりする
- OAI-PMH
  - メタデータを提供する
  - 受け取り側からまとめてメタデータを取得
  - JAIRO Cloudから、DOI取得用にJaLCoが、CiNii Articles用にIRDBがOAI-PMHでデータを取得してる
- OpenURL
  - 書誌情報やIDをURIで送信する
  - リンクリゾルバはこれを受けとって利用できる電子ジャーナル等を返す
  - 電子ジャーナル等からリンクリゾルバにリンクを設定する際にはOpenURLで記述されている
  - ANSI/NISO Z39.88-2004 で規格化
- OpenSearch
  - 検索式と結果をそろえて送受信



32

## データの粒度

- どの単位で書き表しているか。  
そのデータベースではどこまで探せるか。
  - 図書・雑誌のタイトル?
  - 論文・図書の章?
  - 所蔵?
- 例えばNACSIS-CAT/CiNii Books
  - 「書誌情報」と「所蔵」を探ることができる
  - 雑誌論文、目次や章のタイトル、著者では検索できない（ものがほとんど）
- ディスカバリーサービスでは  
この「粒度」を超えて検索できる
  - CiNii Researchでもできるようになる

## データ同士をつなげよう

- データの同定、統合（名寄せ）や連携を効率的に行いたい
- 「IDで結びつける」
  - 図書館目録の世界なら「典拠ファイル」を用意する
  - 相手のデータベース上のIDがわかれば確実にリンク
    - JAIRO CloudとCiNiiの連携はNCIDでリンク
    - OpenURLなどでIDを渡す仕組みが作られている
- IDの例
  - DOI
  - ORCID
  - VIAF
  - (Linked Open Data)

## DOI

- Digital Object Identifierの頭文字
  - コンテンツの電子データに付与される国際的な識別子
  - 論文や図書一冊や章、図表やデータなど様々な単位に設定できる
- コンテンツへの永続的なアクセスを保証
  - 同じ雑誌で出版社等がかわってもアクセスできる
  - 個別のコンテンツに割り振られたID (DOI) とその所在URL情報をペアで保管
  - DOIの問い合わせに対して所在URLを返す
- DOI登録機関 (Registration Agency: RA) が登録を行う
  - CrossRef, DataCite, JaLCなどのRAがある

## ORCID

ORCID ID  <https://orcid.org/0000-0002-5189-1865>

- 研究者その他を一意に識別するコード
  - Open Researcher and Contributor ID
  - 非営利団体 ORCID Inc.が管理する
  - 誰でも無料で登録できる
- 氏名、所属にひも付く
  - 同姓同名の識別に便利
  - 論文の投稿時に入力すると業績も登録される
- 「誰が登録した」かも記録される
  - 自身で業績を登録できる
  - 自身での登録以外に、出版社や所属組織など信頼できる第三者が登録することで情報に信頼性を与える (ここが有償)

## VIAF

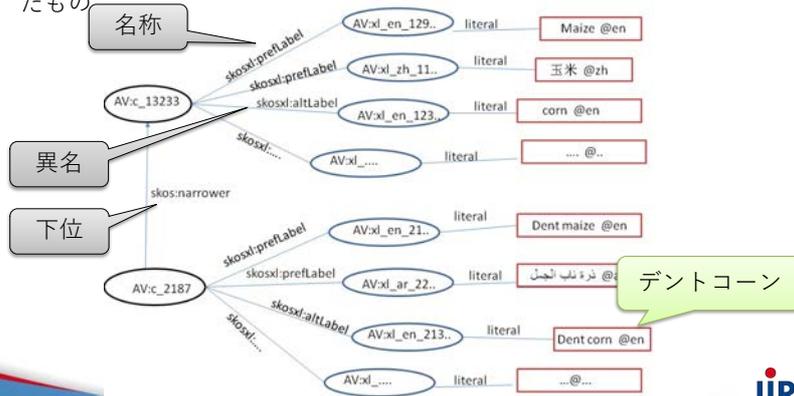
- バーチャル国際典拠ファイル  
(Virtual International Authority File)
- 各国の国立図書館等からの典拠データの提供により、個人、団体といった同一の著者等の典拠データを同定し、名寄せした国際典拠ファイル
- 世界各国の典拠ファイルを相互に結びつける

## Linked Open Data (LOD)

- LODとは
  - 事物をURIで表して相互につなぐ(Linked)
    - そもそもURI=Uniform Resource IdentifierなのでIDの一つ
  - 誰でも自由に利用できるデータ(Open Data)
- 「IDで結びつける」を強化できる
- NDLなどの書誌データや典拠データなどで使われている

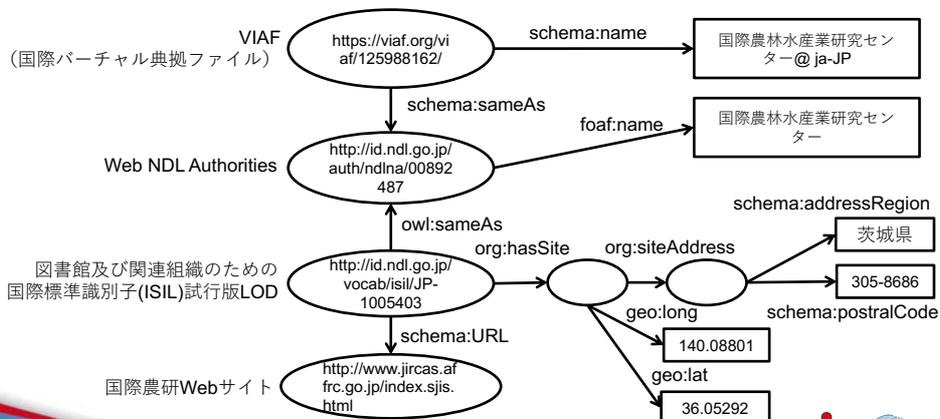
## LODの例

- AGROVOC（農業分野のシソーラス）の表現例
  - 他にも表現例（書き方）があります
- 「AV:c\_13233の英語の名称は”Maize”、別名は”corn”」「下位語はAV:C\_2187」などを表す
- 楕円の中の「AV:C\_13233」やskosxl:prefLabelなどはすべてURIを省略したもの



## LODの例

- 他の情報資源と「つながる」
  - VIAFには組織の緯度経度に関する情報はないが、ISILとつながることでわかるようになる



## まとめ

- 情報サービスがどのように作られているかを意識する
  - データ作成や収集の方法、粒度…
- ディスカバリーサービスだけに頼れない
  - 完全なデータではないかもしれない
  - 使い分けよう
- 図書館員が作成したデータはどこへ
  - OPAC、機関リポジトリだけのものから「みんなが検索できるデータ」に
  - 流通するからこそ、しっかりデータを作成する

## 実習課題から

## 実習課題（おさらい）

各自の大学で導入している有料データベースを調べ、よく使われる3つを実際に検索してみて、その紹介文と取説（検索上の注意点等）をまとめてください。可能であれば収録されている（メタ）データがどこから来ているか、どのように作成されているかも考えて記入してください。導入している有料データベースが少ない場合は、3つの中にCiNii Articles を加えてもかまいません。

（中略）

紹介文はプロバイダーが提供している文そのままではなく、書店のポップ的に、簡潔にまとめてください。（長い説明は利用者に読み飛ばされます。）取説には、どのような情報が得られるか（また得られないか）が分かるよう、収録内容・範囲を必ず記入してください。



43

## 「実際に検索」する際のポイント

- 利用者の要求を考える
  - 特定の論文を入手したい
    - ILLで対応できるか
  - あるテーマ（主題）で調べたい
    - 網羅的かどうか
    - 最新のもの、過去のもの、どこまで対象か
    - 特定のテーマか、主題を問わないか
    - 入手できるものに限るか、文献のリストを作りたいか
    - などなど……



諸条件で選択するデータベースが変わる



44

## 紹介文を書くとしたら

「利用者の要求」を元に情報サービスを整理して、サービスにつなげる（ここでは5W1Hで整理）

### Who（だれが）

- 誰が使えるか。図書館での契約やユーザ登録は必要か。
- 制限はないか（学部生は不可、など）。

### When（いつ）

- いつでも使えるか。サービスの提供時間や開館時間などに限られないか。

### Where（どこで）

- どこで使えるか。学内のみか。自宅（リモート）で使えるか。

### What（なにを）

- 何を検索できるか。できないことは何か。
- 収録対象や範囲は、収録基準はどのようなものか。

### Why（なぜ）

- なぜこのデータベースを使う必要があるか。
- どんな検索の際に使えるか。

### How（どのように）

- どんな検索やリンクの機能があるか。



45

## 余談：いにしえのデータベース検索

- 従量課金制
  - 「サブスク」じゃない！
- 対象データベース、検索式、出力内容を事前に検討する必要
  - 目的に合致しないデータベースは使えない
  - 結果出力一件、さらに表示項目の内容で課金される
  - 大量に出力すると大量の経費
  - 時間がかかると通信費がかさむ



時間、費用など主にコストの制約から  
利用者の要求内容を検討する必要（今でも変わらない）

46

## 補足：より深く知るための情報源



47

## よく講師がチェックする情報源

- カレントアウェアネス・ポータル（国立国会図書館関西館図書館協力課）  
<https://current.ndl.go.jp/car>  
[https://twitter.com/ca\\_tweet](https://twitter.com/ca_tweet)
  - 毎日更新
  - 図書館に関する館種を問わない国内外の情報
- ユサコニュース（ユサコ株式会社）  
<https://www.usaco.co.jp/>
  - 毎月更新
  - 主に国外の学術情報サービスの動向
- STI Updates <学術情報流通ニュース>（科学技術振興機構）  
<https://jipsti.ist.go.jp/johokanri/>
  - 毎日更新
  - JSTの担当者が収集する各国の科学技術情報の生産、管理、流通、活用に関する情報を、速報として簡潔にまとめて紹介



48

地球と食料の未来のために



[www.jircas.go.jp](http://www.jircas.go.jp)