

レポート2の提出期限は6/22(水)
6/17(金) 3-4 時限は美幌博物館の見学。
バスロータリーに停車中のバスに乗車して集合

第9講 データベースとそのネット公開

1. データベースとは

1) コンピュータを用いる

データベース database: コンピューターで、相互に関連するデータを整理・統合し、検索しやすくしたファイル。また、このようなファイルの共用を可能にするシステム。DB (デジタル大辞林)

データベース (英: database, DB) とは、検索や蓄積が容易にできるよう整理された情報の集まり。通常はコンピュータによって実現されたものを指す (ウィキペディア)

データベース 論文、数値、図形その他の情報の集合体であつて、それらの情報を電子計算機を用いて検索することができるように体系的に構成したものをいう (著作権法第2条10の3)

つまりデータベースとはアプリを使ってデータを整理したり、検索したり、並べ替えたりして使いやすくする仕組みといえそうだ。コレクションに対して資料データベース、書籍に対して蔵書データベース、という形で構築される。資料そのものではなく、それを使う仕組みである。

2) アーカイブとの違い

アーカイブ archive は日本では公文書館と訳されてきた。過去のものとなった文書の保管庫である。現実には私文書や写真のアーカイブも存在するため、現在では固有名詞を除きアーカイブとそのまま呼ぶことが多い。インターネットで閲覧可能なデジタルアーカイブはデータベースと混乱しそうであるが、資料の保存庫がアーカイブであり、それを閲覧する仕組みがデータベースである。

使用感の違いは、データベースと名乗るサービスは情報の検索に加え並べ替え (整列、ソート sort) が可能なのに対し、デジタルアーカイブは検索のみという印象である。データベースは目的とするデータとメタデータの両方を操作対象にするが、デジタルアーカイブで用いられているデータベース機能は対象がメタデータに限られている。

以上のことから「電話帳もデータベースの1種」という言い方は不適切、少なくとも不親切である。紙媒体を操作するデータベースは可能であるが、紙媒体それだけを指してデータベースという言い方は無理なことがわかる。「電話帳」がアプリの場合はデータベースだが、文字どおりの冊子体の場合は、やはり台帳である。

3) 身近な例

大学生にとって最も身近なデータベースは大学図書館の蔵書検索システムである。オンラインで使えることからOPAC (おぱっく、おーぱっく Online Public Access Catalog) と呼ぶ。ウィキペディアにはカード方式の蔵書検索に比較したデータベースの優位性や日本語に特有の課題などが簡潔に解説されている。

OPAC - Wikipedia <https://ja.wikipedia.org/wiki/OPAC>

東京農業大学 学術情報センター[オホーツク]

<http://libopac.bioindustry.nodai.ac.jp>

CiNii Books <https://ci.nii.ac.jp/books/?l=ja>

全国の大学図書館OPACを対象にした横断検索システム。学情センター経由で借用するには学情センターOPACトップページ「全国大学図書館を検索」からが便利。借用依頼までスムーズにおこなえる。



コンピュータ以前はカードで蔵書を管理検索した
By Marcus Gossler via Wikimedia (c) Dr. Marcus Gossler
<https://commons.wikimedia.org/wiki/File:Schlagwortkatalog.jpg>

2. 博物館資料の情報化

1) 資料とするための情報化

自然物と自然史資料の違いは、採集情報の有無にある。自然物を「いつ・どこで・だれが」採集したかという最低限かつ最重要な情報を付加することが、自然物を博物館の自然史資料とすることである。人文系の資料も同様で、作品や生産物を資料化するには収集情報を付加することでもある。博物館資料の管理は物体とメタデータの両面が必要である。加えて博物館資料には資料番号（収蔵番号、個体番号）が与えられる。これは学籍番号とおなじ唯一無二の固有番号（ID number. IDはidentityの略）である。歴史ある大規模館の場合、異なる分野で同一番号が重複して用いられている例もあるが、個体の識別は可能となっている。

2) 資料管理のためのメタデータ（＝データのデータ）

①所在情報 博物館資料で保存場所の特定は必須である。保存場所が不明になると資料にアクセスできず、利用も展示も研究も不可能となる。資料や標本が小さく定型の場合は保存の場所も一か所にまとめることも容易で一覧性も高く、行方不明となることは少ない。昆虫標本や腊葉標本（さくようひょうほん、押し葉標本）、絵はがきや行政文書がこれに相当する。反対に形が不定あるいは大型の資料は保存の空間も大きくなるため分散して収蔵することも生じ、時として行方不明になりがちである。民俗資料や鯨類の骨格などがこれに相当する。昆虫や植物の標本は自然物の形はさまざまだが、標本にする段階で標本箱や台紙を用いることで定型化に成功している。実際の館内での所在情報は、部屋（番号）と棚（番号）という組み合わせで実現する。

②利用情報 資料は研究利用やメディアや一般からの撮影、展示などでの利用や貸出があり、資料の保管や管理からこれらの情報の記載も欠かせない。これらの利用は資料の保存状態に影響を与えること、展示に加えてこれらの利用状況が博物館の評価の指標となるため備考でよいのでメタデータに加えたい。

③生活史情報 資料が生体の場合、生活史の情報も必須である。とくに寿命の長い生物では飼育や栽培の過程での怪我や病気、事故などを記録する。これらの記録は次世代以降にも影響を及ぼす可能性もあり、植物園のエキシカータ（交換用種子標本）や動物園ではブリーディングローン（繁殖用の動物の貸出）によって他の館園にも関係する。

3) 公開用データベース

館内で利用する資料データベースは、そのままの形で一般公開することは通常不可能である。それは寄贈者や購入価格、希少種の採集地、関係者の間での共有だけを前提として特記事項などである。寄贈者については資料の来歴として必要な場合、寄贈者や関係者の意向により公開することもある。採集地も普通種や採集が困難な場合は公開しても問題は生じない。しかし希少種や過剰な採集、カメラマンや「映え」ねらいの撮影の集中が予想される場合は非公開とする必要がある。これらの判断は基準を定めて一律に非公開とする、あるいはデータベースに公開非公開の項目を作成して対応するのが現実的である。



上：昆虫標本は統一サイズの標本箱で保存する（北海道大学総合博物館）

中：腊葉標本も統一サイズの台紙に貼って保存する

下：属ごとにカバーして整理（兵庫県立人と自然の博物館）

3. データベースの実例

1) 国立科学博物館 標本・資料データベース <https://www.kahaku.go.jp/research/specimen/>

タイプ標本、地域魚類、魚類写真、音声、文献などさまざまなデータベースのリンクページ。デジタル標本もある。個別のデータベースは実際に使用してレポートしてほしい。

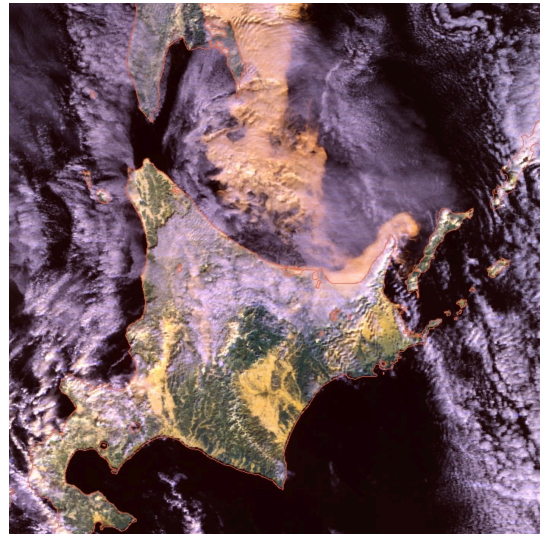
2) YList 植物和名-学名インデックス <http://ylist.info>

日本産植物（緑色植物：コケ、シダ、種子）の和名と学名の対応表。単純な内容に見えるが、学名の変更に対応し、近年急速に再整備が進む分類体系にも対応する。

3) オホーツク海の海水分布拡大画像ページ-JAXA EORC

<http://sharaku.eorc.jaxa.jp/cgi-bin/adeos2/seaice/seaice.cgi?lang=j&mode=large>

日付を指定してその時の衛星画像を表示できる。過去データも検索可能。学情センター入口と同様の画像が得られる。



オホーツク海の海水分布拡大画像ページ-JAXA EORC 2005-2-7の流氷。「高解像度画像へ」をクリックしたもの

研究機関のデジタルデータであるが参考に。

4) アメリカ自然史博物館の脊椎動物学データベース <https://emu-prod.amnh.org/db/emuwebamnh/index.php>

資料検索の例：標本番号がわかっている場合 宮城県石巻市牡鹿町鮎川で収集されたマッコウクジラ

Mammalogyをチェック>Submitをクリック>最上部窓に34872を入力>右のメニューで Catalog Number を選択>Submitをクリック→これで目的の標本 *Physeter catodon* マッコウクジラが抽出された。

結果左端のM-34872をクリック→結果詳細 Specimen Details が表示される(右写真)。

が、採取場所 State に疑問がある Oshima (和歌山県紀伊大島) ではなく鮎川(宮城県石巻市牡鹿町) のはず
このような場合は元データにあたる台帳の画像を確かめる。

Multimedia の画像をクリック>現れた写真をクリックして拡大>34872を調べると Aikawaとあった。

*データ原本である台帳の画像が参照できるのは誤りを正すために/発見するために重要

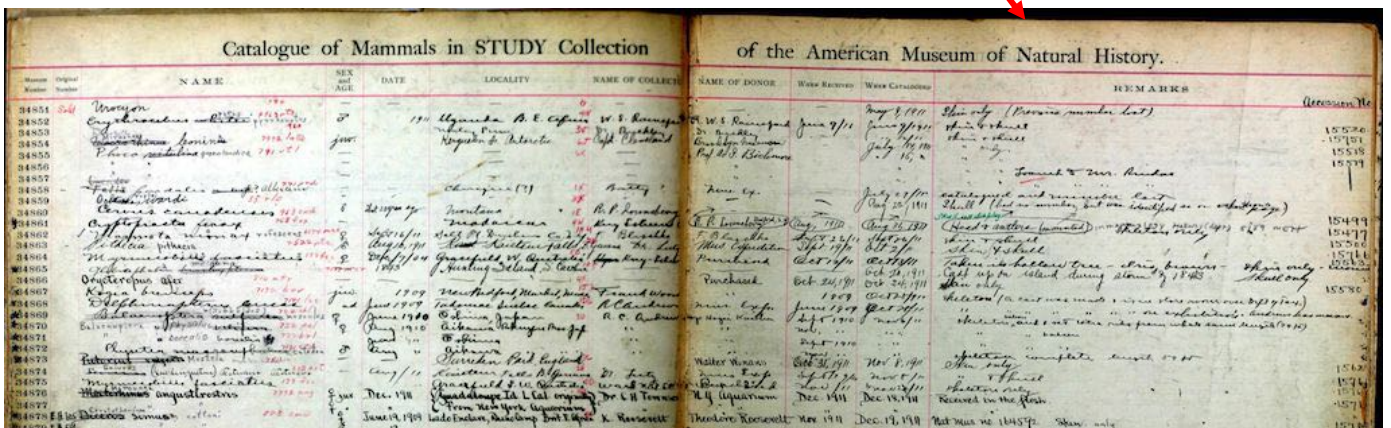
アメリカ自然史博物館の脊椎動物学データベースの検索結果 M-34872

Specimen Details

Record number 1 of 1.

M-34872: *Physeter catodon*: Male: Japan: Oshima: Organism

Catalog Number	M-34872				
Current Identification	Physeter catodon				
Taxonomy	Order	Family	Subfamily	Genus	Species
	Cetacea	Physeteridae		Physeter	
Country	Japan				
State	Oshima				
County					
Specific Locality					
Lat/Long	Latitude				Longitude
Collection Date	01 Aug 1910				
Collector(s)	Roy C. Andrews				
Collector's Number					
Sex	Male				
Type Status					
Preps Available	Prep Type	Part Prepared	Prep Description		
	Skeleton	Complete Post Cranial	Disarticulated		
	Skull	Cranium and Mandible	Not Mounted		
Multimedia					



アメリカ自然史博物館の資料データベースでは台帳に遡ってメタデータの確認作業が可能

4) JBIF: 日本生物多様性情報イニシアチブ <http://www.gbif.jp/v2/>

生物の出現情報（実際には観察情報）や標本を世界規模で収集し、統合データベースにして運用する The Global Biodiversity Information Facility (GBIF) の日本語、日本窓口。GBIFはOECDの参加政府間の覚書により2001年に発足したものの。その前段には1992年に採択、翌1993年に発効した生物多様性条約（Convention on Biological Diversity CBD）があった。そもそもCBD以前に生物多様性という言葉は一般には使われていなかった。



GBIFとは何か? <https://www.gbif.org/ja/what-is-gbif>

地球規模生物多様性情報機構GBIFの動きと役割 pdf 905 KB https://www.jstage.jst.go.jp/article/seitai/66/1/66_209/_pdf

細谷剛 (2016) 日本生態学会誌, 66: 209–214. 科博研究部の研究員による解説

JBIFは和名での検索が可能であるが種名を特定する必要がある。「コケ」で検索すると和名の一覧がずらりと並び結果となり、そこから和名一つ一つをクリックして初めてデータに行き着く。データの提供は熱心な研究者や学芸員が在籍している機関が牽引し、110機関386データセットと発展途上であるが相当程度集まってきている。

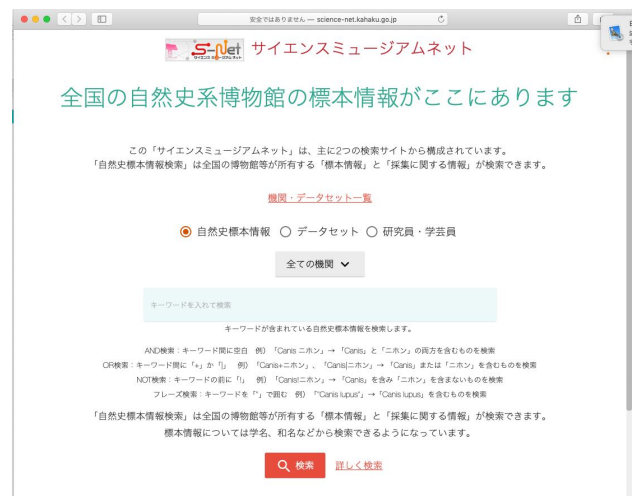
6回目の授業でメタデータの標準化として Dublin Core を紹介した。これを自然史資料に特化させたものに Darwin Core があり、GBIFも採用している。

データ形式 | JBIF: 日本生物多様性情報イニシアチブ http://www.gbif.jp/v2/dataset/data_format/index.html

4. サイエンスミュージアムネット (S-net)

1) S-netとは <http://science-net.kahaku.go.jp>

S-netは「全国の科学系博物館の協力の基に国立科学博物館が運営するポータルサイト」（公式サイトの説明）。国内の博物館が所有する自然史標本と学芸員（研究員含む）の2つの情報を対象にした横断検索システムである。「自然史標本情報検索」は標本情報と採集情報が、「研究員・学芸員検索」は博物館や研究機関の研究職員が検索できる。



参加は任意であるが主要な自然史博物館が多数参加して

いる。網走近郊では美幌博物館、道内では他に北海道教育大学旭川校、小樽市総合博物館、霧多布湿原センター、北海道大学総合博物館、釧路市立博物館、帯広百年記念館、北海道大学総合博物館水産科学館（函館）が参加している。全国的にもJBIFよりも参加機関が多い。

2) 目的と特徴

目的は日本語で記載された自然史資料の横断検索システムの構築と運営である。科博や協力者の伝道師が国内を回り説明会を開いたことで参加機関が増加しJBIFよりも充実したデータを揃える。また、データの標準化のアプリとマニュアルが行き届いており、コンピュータに少し詳しくればデータの変換作業も容易。学芸員が検索できるのも画期的である。ICOM日本委員会や日本博物館協会ではウェブでもリアルでもこのようなサービスはおこなわれていない。データの検索はデータセット（コレクション）で可能なほか、「コケ」といった曖昧な語でもデータが直接抽出され直感的で使いやすいと感じる。

3) 活動実績・資料 <http://science-net.kahaku.go.jp/app/page/activity.html>

「活動実績・資料」というページでは、パンフレットや使い方の説明のほかデータベースを用いた研究などの活用事例の報告集、過去の研究会や説明会の資料、一番下には2015年に実施された「日本国内における自然史標本資料の電子化状況アンケート調査結果」にリンクしている。リンク先には pdf が多いので注意。

S-netを知るのが目的であれば上部にある「S-netの使い方」のリンク先から見て行くのがよいだろう。

4) データの標準化

横断検索や統合データベースの構築にはデータ項目や名称、単位、さらにはデータの階層化などの標準化が必須である。S-netでは標準化のためのアプリとマニュアルが整備され、インターネットからダウンロード可能である。アプリはエクセルを応用したもので実質的に博物館関係者なら誰もが使用可能なものとなっている。

S-netも標準化には Darwin Core を大きく取り入れている。ただし日付など一部は独自様式である。

アプリでは、たとえば地名は一続きの住所を都道府県、市町村、詳細に切り分けすることや、日本語を英語に変換

する作業が自動的におこなわれる。また住所が不明だったり複数の呼称が存在する海岸地名については、「日本沿岸地名表」（海上保安庁水路部 1982）をエクセルにした「日本沿岸地名辞書」を用いている。

「参加機関・参加検討中の機関の方へ」は日本沿岸地名辞書やデータに関する説明をリンク付きで解説している。

http://science-net.kahaku.go.jp/app/page/tool_download.html

5) 参考資料

ここまで説明してきたが科博の研究者による解説やマニュアルが公開されており、それを参照してほしい。

S-Net自然史情報を構成する標本データとその背景文化 <http://www.gbif.jp/v2/pdf/Fukuda.pdf> [media2022_9-2.pdf](#)

サイエンスミュージアムネットデータ変換ツール利用者向けマニュアル [media2022_9-3.pdf](#)

http://science-net.kahaku.go.jp/contents/resource/SNet31_20180514_DataConvManual_20180511.pdf 自動リンクしないのでコピペが必要

	A	B	C	D	E
1	読み	地名	都道府県	緯度(日本地測)	経度(日本地測)
2	Aba Sima	阿波島	広島県	34 19	132 57
3	Abasiri Kawa	網走川	北海道	44 01	144 17
4	Abasiri Ko	網走湖	北海道	43 58	144 11
5	Abasiri Ko	網走港	北海道	44 01	144 17
6	Abasiri Wan	網走湾	北海道	44 06	144 46
7	Aba Wan	網走湾	長崎県	32 45	129 58
8	Abe Kawa	安部川	静岡県	34 56	138 24
9	Abira Kawa	安平川	北海道	42 37	141 44
10	Abose	アボセ	和歌山県	33 43	135 21
11	Abosi Ku	網干区	兵庫県	34 45	134 36
12	Abukuma Kawa	阿武隈川	宮城県	38 03	140 55
13	Abumi Hana	鑿鼻	岩手県	39 19	141 59
14	Abuoru Sima	安部王留島	沖縄県	26 32	128 06
15	Aburame Saki	油目埼	青森県	40 59	140 52
16	Abura-no-Oki-no-Isi	油ノ沖ノ石	愛媛県	34 16	133 12

S-net で用いる「日本沿岸地名辞書」

No	項目名	必須	値
システム管理情報			
1.	管理 ID		システム管理用項目
2.	登録日時		システム管理用項目
3.	更新日時		システム管理用項目
4.	データセット管理番号		システム管理用項目
5.	グローバルユニーク番号		システム管理用項目
基本情報			
6.	データ登録日時		
7.	GBIF 公開フラグ	○	0: GBIF で公開しない, 1: G 定値)
8.	レコード種別	○	PreservedSpecimen, Foss LivingSpecimen, HumanO MachineObservation, Mat Occurrence のいずれか
9.	機関名	○	機関名の英文名称
10.	機関名 (日本語)	○	機関名の和文名称
11.	機関コード	○	標本の機関コード (例: TNS、)
12.	コレクションコード	○	標本のコレクションコード (例: ショーンコードが未設定の機関では)
13.	カタログ番号	○	標本番号
オカレンス情報			
14.	採集者番号		採集者によるオリジナルの標本
15.	オカレンス備考		標本等に関する補足説明 (例 the road)
16.	オカレンス備考 (日本語)		標本等に関する補足説明 (例)
17.	性別		例: male, female

「変換ツール利用者向けマニュアルより」データ項目と必須の有無、内容の一覧

③ 抽出分割(日本語)をクリックすると、順に分割された結果が入ります。「日本」は補完されます。

標準データ事前整形支援ツール [互換モード]

住所を含む列

抽出分割(日本語)

抽出分割(英語)

All Clear

変換件数(1-65534)

100

Topに戻る

「参加機関・参加検討中の機関の方へ」最下段の右側「住所の切り分け」より http://science-net.kahaku.go.jp/contents/hint/RMB21_separateaddress.pdf