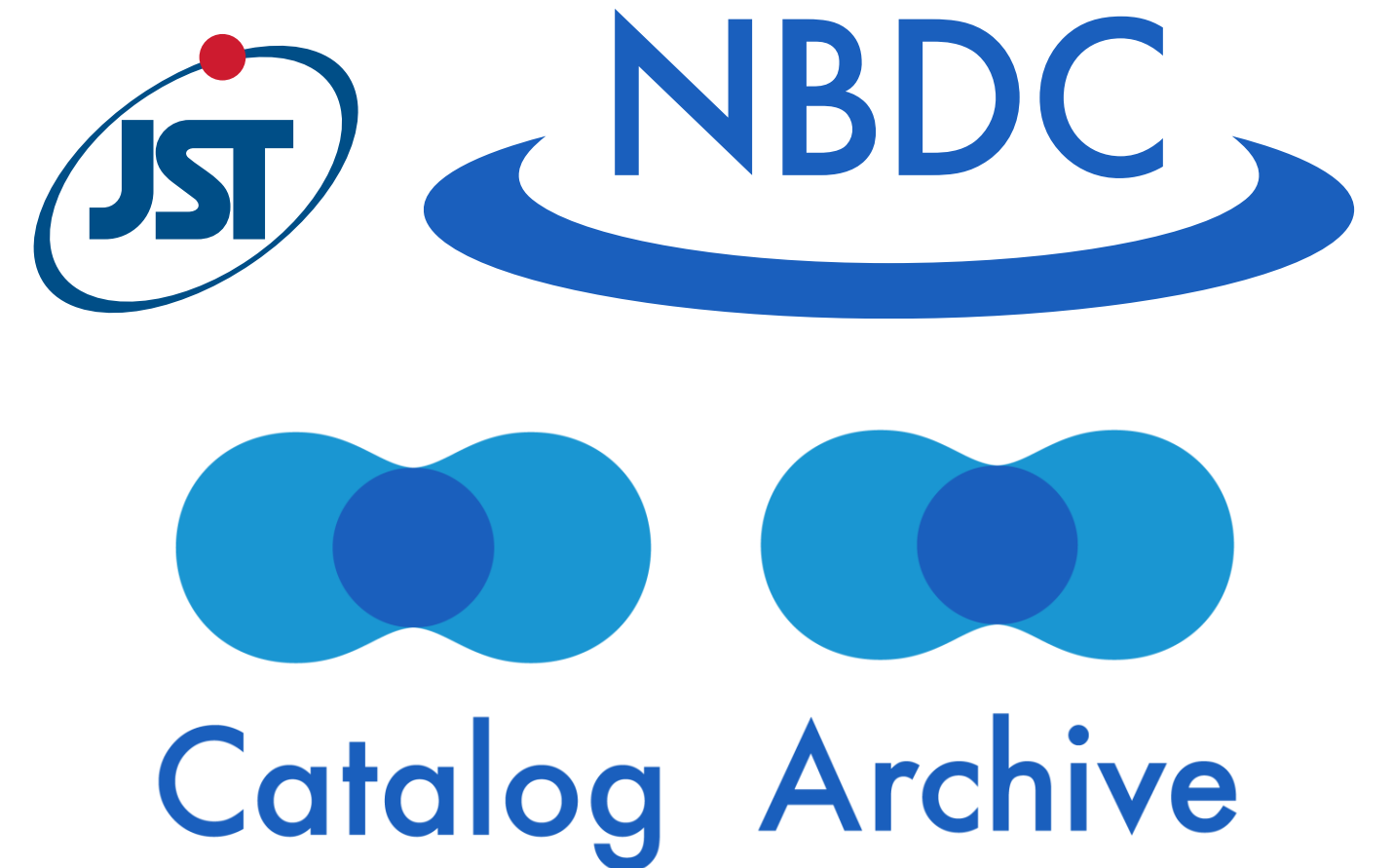


Integbio データベースカタログ・生命科学系データベースアーカイブの紹介

井手隆広（科学技術振興機構・情報基盤事業部・NBDC事業推進室）



要旨

科学技術振興機構（JST）情報基盤事業部NBDC事業推進室（NBDC）は公募型の研究費制度「統合化推進プログラム」を実施する他に、生命科学系のデータベースを探る、預ける、検索するサービスを提供している。これらのサービスは2010年代、増え続けるデータベースに係る要望として内閣府の統合データベース（推進）タスクフォースの示す方針を踏まえて構築された。当時の要望は「どんなデータベースがどこにあるのか?」、「データベースに含まれるファイル群を取得できないか?」、「複数のデータベース内を横断的に検索できないか?」であり、その回答としてデータベースの所在情報を集めた「カタログ」、データセットの置き場所としての「アーカイブ」、データベースの一括検索を実現する「横断検索」が機能している。本ポスターでは、カタログとアーカイブについて紹介する。

#データベース、#カタログ、#アーカイブ、#FAIR原則

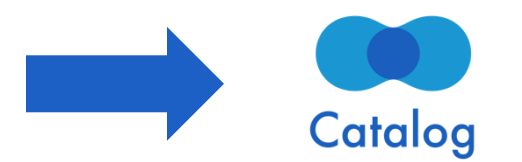
NBDCのページ



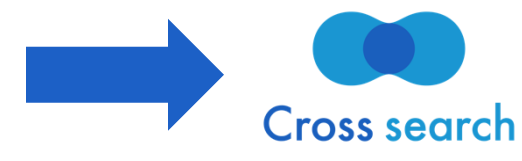
データベースに対する要望

2010年代に数多くの生命科学系データベースが誕生した（している）

- どんなDBがどこ(URLは?)にあるのか?
 - 欲しいデータは既に誰かが持っているかも
 - 公開しているかも
- データセットを取得できないか?
 - インターフェイスにとらわれず自由に使いたい
 - 手持ちのデータと組み合わせてみたい



- 複数のDB内を横断的に検索できないか?
 - DBごとに訪れて検索するのは手間
 - 文献や特許など異なるDBを一括検索できれば



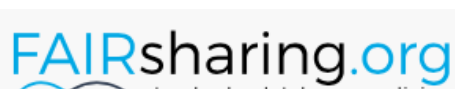
(AJACS 資料より https://github.com/AJACS-training/AJACS93/bioib/main/01_minowa/AJACS93_01_minowa.pdf)

Integbioデータベースカタログ



(文科省) (厚労省) (農水省) (経産省)

- 文部科学省、厚生労働省、農林水産省、経済産業省の4省による生命科学系データベース統合の一環として提供
- 2572件(日本語)、2170件(英語)のデータベースの情報を収載(2024年8月末)
 - 所在情報(国内外)、説明、生物種など
- 国内の公的研究費で作成されたデータベースを網羅し国内外に発信
 - 独自に研究報告書(KAKEN, GRANTS, AMEDfind)の調査
- Nucleic Acids Research Database Issueから国外のデータベースを収録
- 国際的な生命科学系データベース情報を収集・公開しているFAIRsharingと連携データベース情報を交換(2017年から)
- 日英サイトを用意している
- ツールは対象外



生命科学系データベースアーカイブ

<https://dbarchive.biosciencedbc.jp/index.html>
「LSDB archive」で検索

- 2009年
 - 「ライフサイエンス分野における統合データベースプロジェクト(統合DBプロジェクト)」の一環として、ライフサイエンス統合データベースセンター(DBCLS)でサービス開始
- 現在
 - 155データセットを収録(2024/8月末時点)



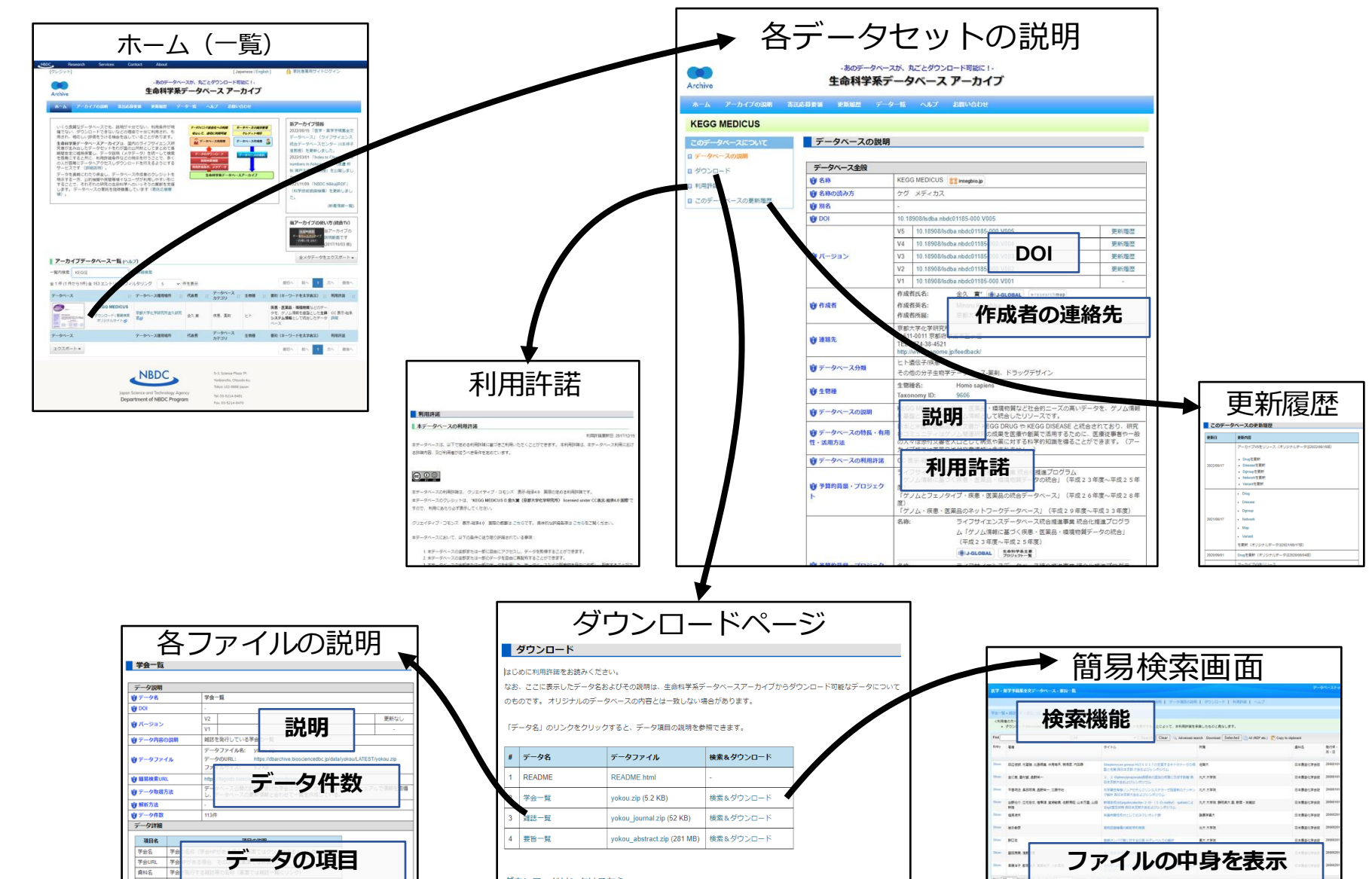
アーカイブ化のメリット

- 長期間にわたる安定的なデータの維持・保管
 - 寄託後は、NBDC内部でメンテナンス
 - 遠隔地にバックアップを保管するBCP対策
- 利用者に見つかりやすく
 - Integbioデータベースカタログへの登録
 - 簡易検索に登録されたデータは生命科学データベース横断検索から検索可能
- 利用者に使いやすい
 - 全てのデータセットで統一的なインターフェイスを用いることで中身を把握しやすくなる
 - 利用許諾を明確にして悩まないように

ページと機能



各ページの構成



最近注目のデータセット

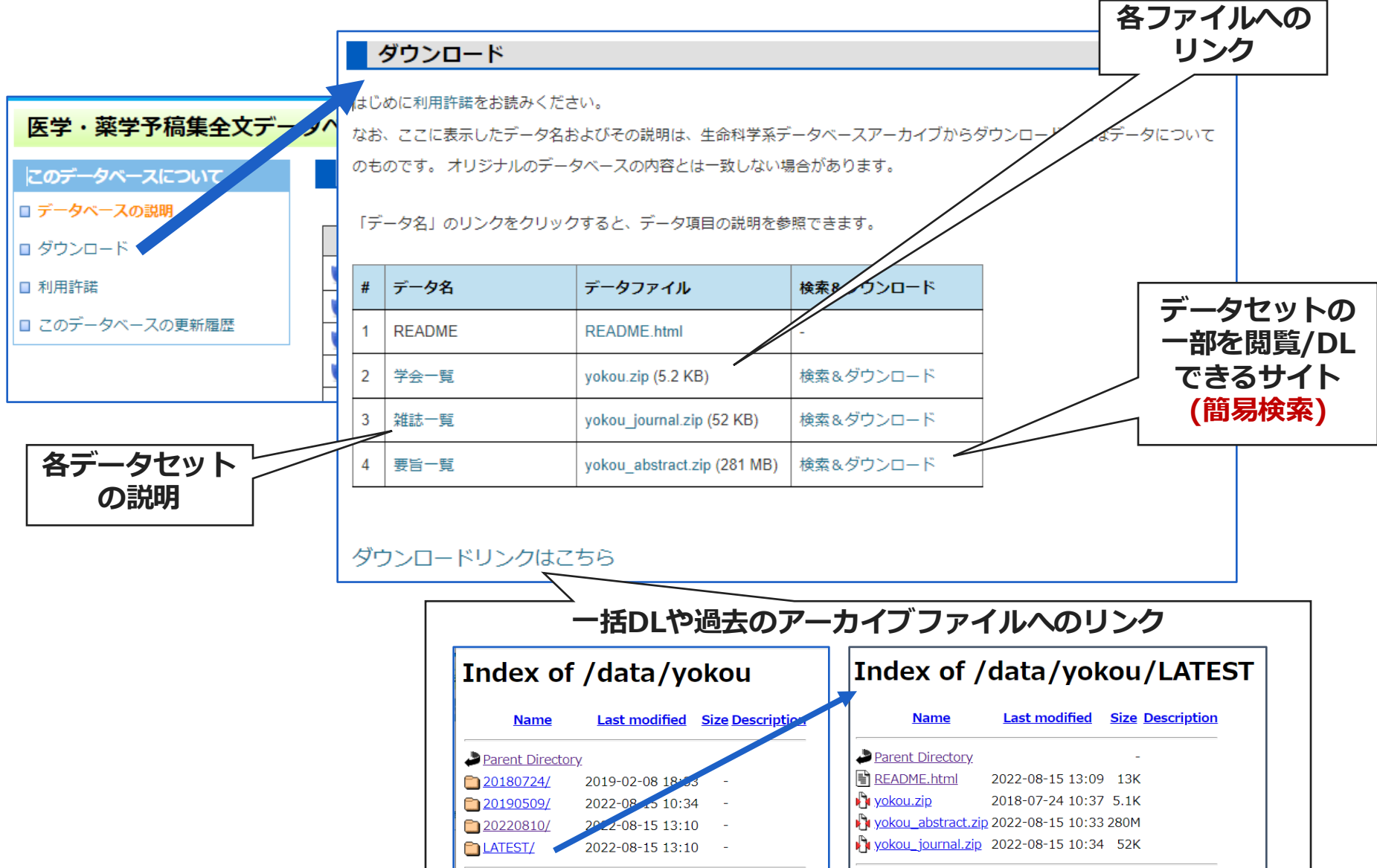
順位	データベース	データベース運用機関	代表者	データベースカテゴリ	生物種	要約	利用許諾
1	Open TG-GATES 病理写真データベース	ダウンロード 簡易検索 オリジナルサイト	東京大学先端科学技術研究センター	トキシコゲノミクスデータベース	ラット	160の化合物を用いて作成した動物試験で得られた肝臓・腎臓の病理標本を高解像度で提供しています	CC表示-継承
2	KEGG MEDICUS	ダウンロード 簡易検索 オリジナルサイト	金久 賢	疾患、薬物	ヒト	疾患、薬物、遺伝子などのデータ、遺伝子発現データなどからなる生命科学系データベース	CC表示-継承
3	Togo Picture Gallery	ダウンロード 簡易検索 オリジナルサイト	小野 浩	画像コレクション	-	生命科学分野の誰でも自由に閲覧・利用できる動物実験のコレクション	CC表示
4	Open TG-GATES	ダウンロード 簡易検索 オリジナルサイト	東京大学先端科学技術研究センター	トキシコゲノミクスデータベース	ヒト、ラット	170の化合物をラット肝臓・腎臓を用いて作成した動物試験で得られた肝臓・腎臓の病理標本を高解像度で提供しています	CC表示-継承
5	Index to Chromosome numbers in Asteraceae	ダウンロード 簡易検索	渡邊 明	植物	植物	キク目に含まれる13科全部の染色体番号をまとめたデータベース	CC表示-継承

各レコードのページ

- 利用許諾
- 連絡先
- 論文へのリンク
- FAIRsharingへのリンク



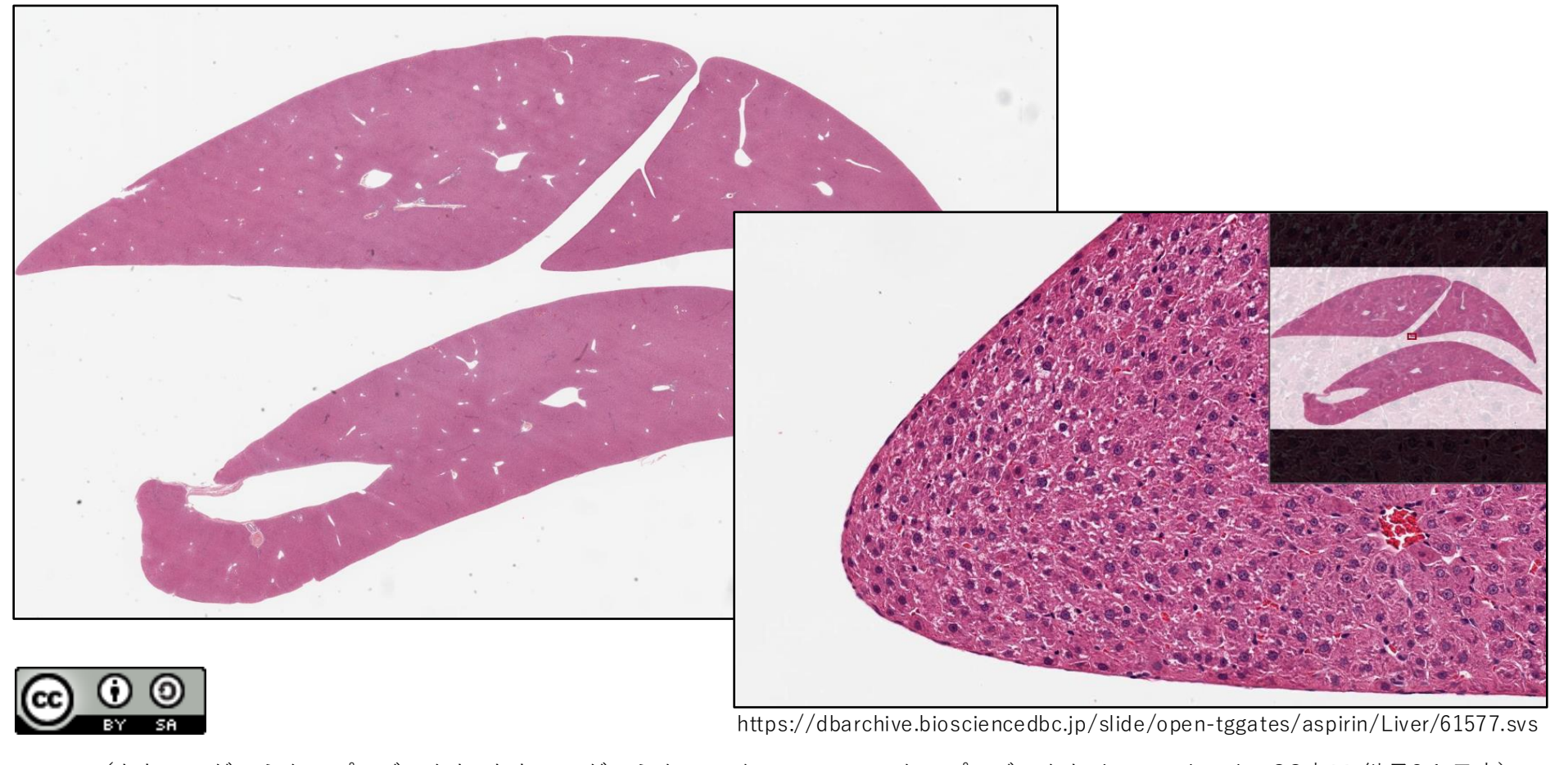
ダウンロードページ



(紹介) Open TG-GATES 病理写真 DB

- Open TG-GATESに登録された160の化合物について、動物試験で作成した肝臓・腎臓の病理標本を高解像度で提供しています

(アスピリン・ラット・投与後24時間・1000 mg/kg)



カタログのデータはDL可能

- RDFデータもあり@RDFポータル



データベースIDを別、URL、運用機関名(ID-U-LOCAL ID)、運用機関名、運用機関所在地、説明、生物種、タグ(特記タグ(データの種別))、論文等(PubMed ID)、言語、移動状況、一括ダウンロードデータのデータ説明、データベースの利用許諾へのリンク、データベースの使い方、データベースの連絡先、LSDBアーカイブへのリンク、MEDALSデータベースへのリンク、統合TVへのリンク、FAIRsharingへのリンク、類似データベース、レコード管理者、レコードの由来、データ取得日、レコード公開日、レコード最終更新日

簡易検索ページ

- csv形式のファイルについては 中身を閲覧可能なサービスを提供
- データの部分的な取得も可能になっている



今後

- 長期安定運用に向けたユーザーインターフェイスの構築
 - 仮想環境の変更に伴う、内部の更新が必要となった
 - 長期運用を見据えメンテナンス性のよいシステムの構築を目指す

カタログ・アーカイブサービスは以下の方々協力のもと運営されております(敬称略・順不同)

建石由佳 (NBDC)、信定知江 (NBDC)、宮崎敦子、畠中秀樹 (DBCLS)

Licensed under a Creative Commons 表示4.0国際ライセンス
©2024 井手 隆広 (科学技術振興機構 情報基盤事業部)

