

人文学のオープンデータ事例

東京大学史料編纂所 久留島典子

事例1

国立国会図書館 NDL

古典籍資料（貴重書等）
(97887点)、憲政資料(3981点)・・・

図書資料に準じた記載情報
→種々の資料に最適化した
情報付加の必要

<http://dl.ndl.go.jp/>



事例 2 - 1

国立公文書館

<https://www.digital.archives.go.jp/>

階層構造を示すなど、資料の特性に応じた情報を持つ



The screenshot shows the search results for '公文録' (Kōmonroku) on the National Archives of Japan Digital Archive website. The page is in Japanese and features a green header with the site name and navigation buttons. A left sidebar contains a tree view of document categories. The main content area displays a list of search results, with the first result being '公文録 - 明治10年' (Kōmonroku - Meiji 10). Below the list, there are search filters and a table of results.

No.	検索結果
1	公文録 - 明治10年
2	公文録 - 明治11年
3	公文録 - 明治12年
4	公文録 - 明治13年
5	公文録 - 明治14年

事例 2 - 2

アジア歴史資料センター <https://www.jacar.go.jp/>



2019/6/9

事例 3

京都府立京都学・歴彩館

<http://hyakugo.kyoto.jp/>

『東寺百合文書目録』(1976-1979年刊行)のデータをもとに画像公開

原本24,067通の保管と
目録・画像データの公開



2019/6/9

SAT大正新脩大藏經テキストデータベース2018版 (SAT 2018)

The screenshot displays the SAT 2018 web application interface. At the top, there is a search bar and navigation links. Below the search bar, there are tabs for 'はじめに', '検索結果', '原文画像', '日本語訳', '英語訳', and 'BIF Linking'. The main content area is split into two columns. The left column shows a scanned manuscript page with a dark background and white text. The right column shows the corresponding digital text transcription in Japanese. The interface includes various controls for zooming, panning, and switching between the original image and the text.

検索履歴

最近のあなたの閲覧・検索履歴

この閲覧・検索履歴は、Local Storageというブラウザの機能を利用して、利用者の環境のWebブラウザに記録しております（サーバー側には残っておりません）。なお、GoogleやFirefox等でLocal StorageやCookie等を食った設定ファイルを持つ複数ブラウザで閲覧している場合、どのブラウザからでもこの履歴が見えるかもしれません。

履歴を削除

すべての履歴を削除

1件 閲覧: 1000 及び 1000 年 (2018/01/01 - 現在)

事例 5

情報・システム研究機構人文学オープンデータ共同利用センター

<http://codh.rois.ac.jp/>

データセット一覧

データセットの提供



日本古典データセット

歴史的典拠データベースにおいてデジタル化された古典籍のうち、主に国文学研究館蔵本を対象に、画像データと音読データセットで公開しています。さらに一部の古典籍には作品紹介や著者プロフィール、タグ情報なども付与しています。



江戸料理レシピデータセット

日本古典籍データセットに含まれる江戸の料理本を対象に、江戸の料理文化に関するデータとして、書籍・地域・部材・レシピ化という作業を加えたレシピデータを提供します。



日本古典縦くずし字データセット

日本古典籍データセットで公開するデジタル化された古典籍を対象に、縦向きテキストの特殊処理で読みやすくし字の並び出し順に字形と書体情報などを、機械学習



KMNISTデータセット

日本古典縦くずし字データセットを元に、機械学習研究で有名なMNISTデータセット互換のくずし字データセットKMNI1を作成しました。目印に応じて、Mandarin-MNIST、Mandarin-10、Mandarin-Kanjiの3種類のデータ

事例 6

東京大学史料編纂所

<https://www.hi.u-tokyo.ac.jp/index-j.html>

- ・ 他機関所蔵資料
の目録公開
リンクして画像公開
- ・ 種々のデータ
ベース
→整理の必要



木簡画像DB（奈良文化財研究所）と 電子くずし字字典（東大史料）との連携

奈良文化財研究所 奈良文化財研究所

『電子くずし字字典データベース』『木簡画像データベース・木簡字典』連携検索



『電子くずし字字典データベース』『木簡画像データベース・木簡字典』連携検索へようこそ

このサイトでは、奈良文化財研究所と奈良文化財研究所、電子くずし字字典データベースとして電子くずし字辞典、木簡画像データベースとして、電子くずし字辞典データベースを構築しました。両者を連携して検索できるようにし、検索結果を共有できるようにします。検索に活用いただけるよう、連携検索を行います。

電子くずし字辞典データベースは、奈良文化財研究所が所蔵する木簡・木簡画像データベースと連携して検索できるようにしています。電子くずし字辞典データベースは、奈良文化財研究所が所蔵する木簡・木簡画像データベースと連携して検索できるようにしています。

[連携検索へようこそ](#)

[電子くずし字辞典データベース](#)

お問い合わせ先
奈良文化財研究所 調査課 調査係
〒746-8501 奈良県奈良市大宮町1-1-1
TEL: 074-367-2111 FAX: 074-367-2112
E-MAIL: kenkyu@nara.go.jp

Copyright © 2019 Nara National Institute of Cultural Heritage. All rights reserved.

奈良文化財研究所 奈良文化財研究所

『電子くずし字字典データベース』『木簡画像データベース・木簡字典』連携検索



検索文字「愛」

電子くずし字データベース | 奈良文化財研究所 | 電子くずし字辞典データベース

愛	愛	愛	愛	愛
奈良文化財研究所	奈良文化財研究所	奈良文化財研究所	奈良文化財研究所	奈良文化財研究所

検索文字「愛」 | 奈良文化財研究所 | 電子くずし字辞典データベース



奈良文化財研究所 調査課 調査係

日本における人文学系データ公開の特徴と課題 1

• 資料

所蔵機関による公開が中心 事例 1, 2, 3

画像が先行、全文データは限られている

テキスト作成の努力→データセット（事例 5）

→電子辞典（事例 6）

• 研究成果

オープンな公開は機関リポジトリ中心

各機関、IRD (<https://irdb.nii.ac.jp/>) 収載などは少数

雑誌論文・著書は課題

（目録はCinii (<https://ci.nii.ac.jp/>) でほぼオープンな公開実現)

辞典類等の一部は商業的に電子的公開 ジャパンナレッジ Lib

日本における人文学系データ公開の特徴と課題 2

- 研究経過の公開？ 公開以前の保存・記録？



- 目録・画像・テキストに関連知識・文献等がリンク
データを分析できるデータベース 事例4 分野が限定

【参考】海外の人文学系研究成果のオープン化事例

Open Library of Humanities. (OLH) <https://www.openlibhums.org/>

2013年に英国・米国の大学研究者3名が立ち上げ、APCではなく、図書館コンソーシアムからの出資金で運営。2017年現在14種のオープンアクセスジャーナルを発行。900超論文掲載

Knowledge Unlatched (KU) <http://www.knowledgeunlatched.org/>

2012年に英国で設立、図書館からの出資金で、人文社会科学系学術書のオープンアクセス化に取り組む

*参考文献：東京大学附属図書館編『オープンアクセスハンドブック』2017年

<http://hdl.handle.net/2261/72694>