

学術情報を見つける、集める、使ってみる

人文情報における課題と未来

関野 樹
(国際日本文化研究センター)

どのような人文系の情報があるのか

国際日本文化研究センター(日文研)では

史資料の画像、翻刻データ

- 古記録、文書、書簡、
- 和歌、俳句
- 新聞、裁判記録

絵画、写真、図像、絵巻、浮世絵

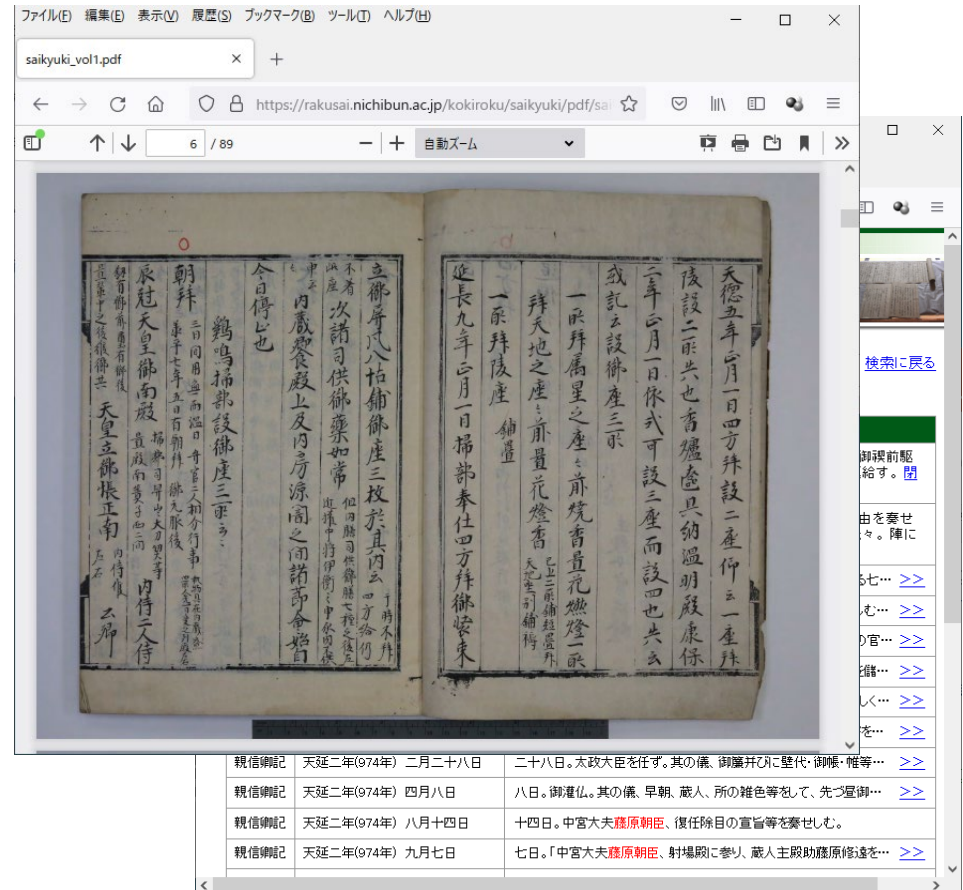
- 妖怪、艶本
- 地図、絵図
- 鯰絵

音楽

- 浪曲

語彙

- 季語



親信御記	天延二年(974年) 二月二十八日	二十八日。太政大臣を任ず。其の儀、御簾並びに壁代・御帳・帷等…
親信御記	天延二年(974年) 四月八日	八日。御灌仏。其の儀、早朝、藏人、所の雑色等をして、先づ皇御…
親信御記	天延二年(974年) 八月十四日	十四日。中宮大夫藤原朝臣、復任除目の宣旨等を奏せしむ。
親信御記	天延二年(974年) 九月七日	七日。「中宮大夫藤原朝臣、射場殿に参り、藏人主殿助藤原修達を…

摂関期古記録

どのような人文系の情報があるのか

国際日本文化研究センター(日文研)では

史資料の画像、翻刻データ

- 古記録、文書、書簡、
- 和歌、俳句
- 新聞、裁判記録

絵画、写真、図像、絵巻、浮世絵

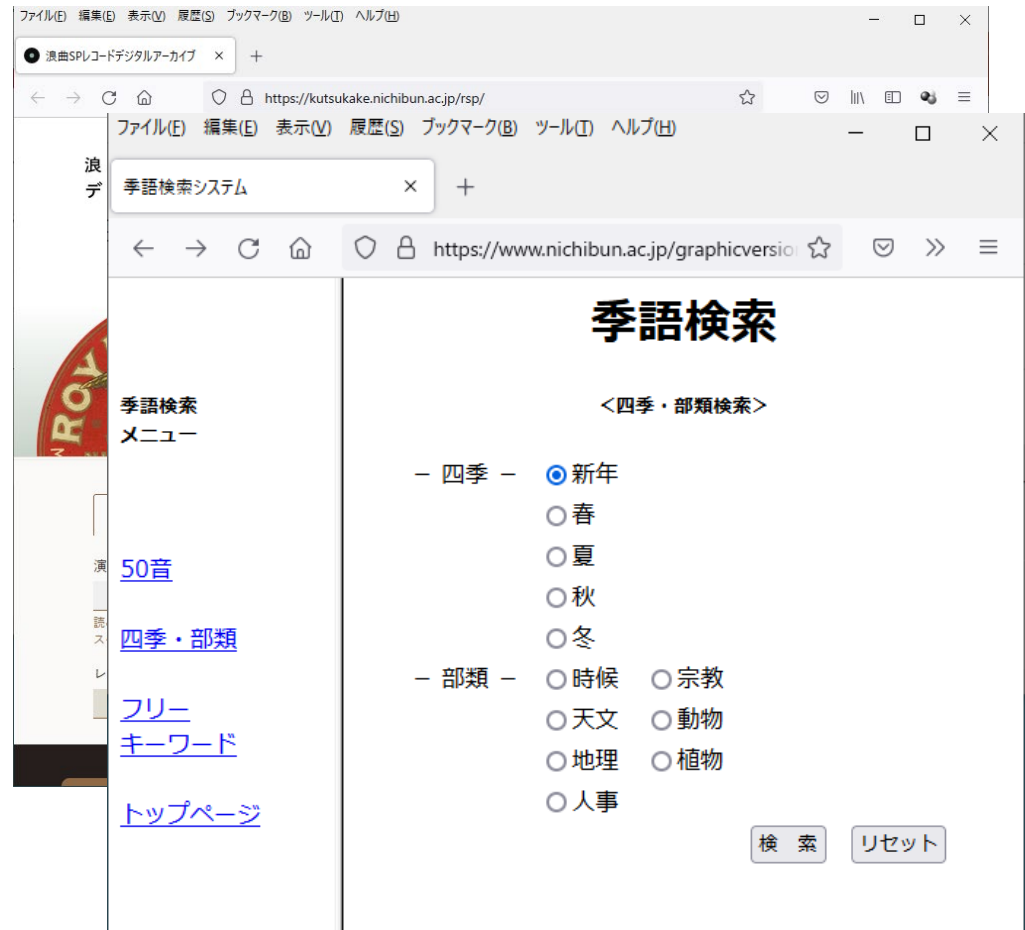
- 妖怪、艶本
- 地図、絵図
- 鯰絵

音楽

- 浪曲

語彙

- 季語



季語検索

どのような人文系の情報があるのか

人間文化研究機構に広げると

目録

- 図書
- 館蔵資料
- 研究者の資料、著作

調査記録

- 発掘記録
- 映像
- 音声(聞き取りなど)
- 調査データ

コーパス

書き言葉や話し言葉に
研究用の情報を付加したもの

梅棹忠夫アーカイブズ (全リスト)

[戻る](#)

資料番号	資料名	形態	該当年	点数
1	著作物 (初版)	現 物	1934-	4,881
		書誌データ	1934-	約7,000
2	著作物 (増刷)	現物	1950-	698
3	著作物の二次利用	現物	1978-	115
4	引紹批言録	A4判 ニつ折りフォルダー	1951-	約8,000
5	編集用資料	A4判 オープンファイル、ニつ折りフォルダー		約1,000
6	一件ファイル	A4判 オープンファイル	1932-	約8,800
7	新聞きりぬき	A4判のクラフト紙に貼付		約5,000
8	現地調査ファイル	A4判 オープンファイル	1932-1997	377
9	民博準備室資料	A4判 ニつ折りフォルダー	1973-1974	約100
10	民博ファイル	A4判 オープンファイル	1973-1993	857
11	フィールド・ノート	冊子	1932-1974	約160
12	日録	冊子	1931	45-
13	ノート	冊子	1936-1943	103

梅棹 忠夫アーカイブズ (国立民族学博物館)

<http://nmearch.minpaku.ac.jp/umesao-archives/index.html>

いち)

『研究会』

人文系の情報の特徴

多種多様である

- 対象が多種多様
史資料、絵画、地図、調査データ、写真・・・
- 発信元が多種多様
大学や研究所以外にも、博物館、美術館、文書館、図書館が重要
- 整理方法が多種多様
メディア別、対象別、コレクション別、・・・
- 発信方法が多種多様
画像同じ史料でも、画像、全文テキスト、断片、・・・

資料を一意に識別する仕組みが不十分

ISBNやdoiのような仕組みが整備されていない

情報の中身

- 専門性が高い
- 閲覧前提での作りになっている
- 画像(映像・音声なども)が多い
- 閲覧制限(個人情報保護など)



情報を見つけるのが
難しい

情報を使うのが
難しい

検索の入りを絞る ― データベースの横断検索

nihuINT

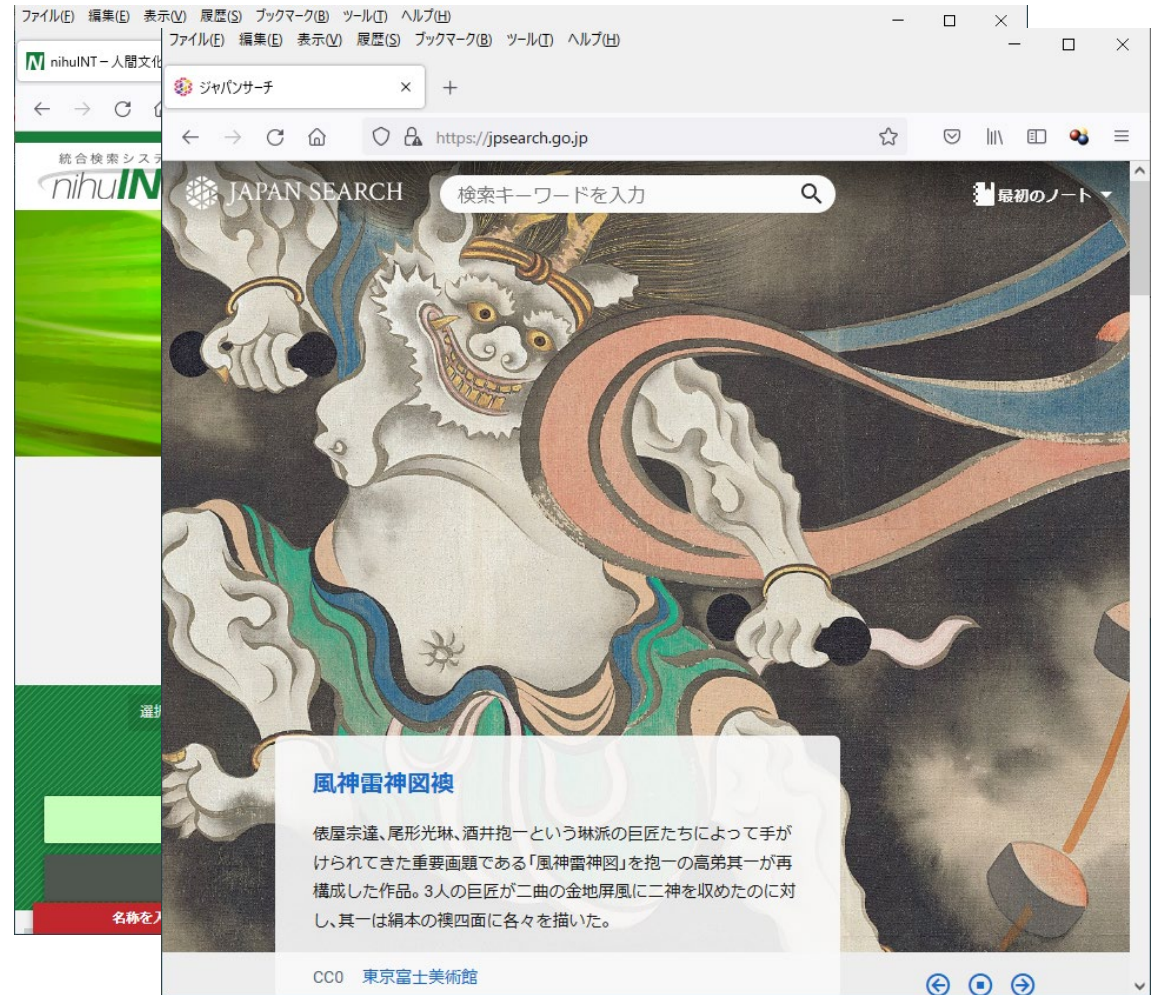
- 人間文化研究機構
- 同機構内の機関＋α

Japan Search

- デジタルアーカイブジャパン推進委員会
- 70機関(2021年10月11日現在) 図書館、大学、政府機関など 人間文化研究機構の機関も含む
- 画像データが豊富

課題

- 欲しくない情報も検索される
- 専門的な検索ができない

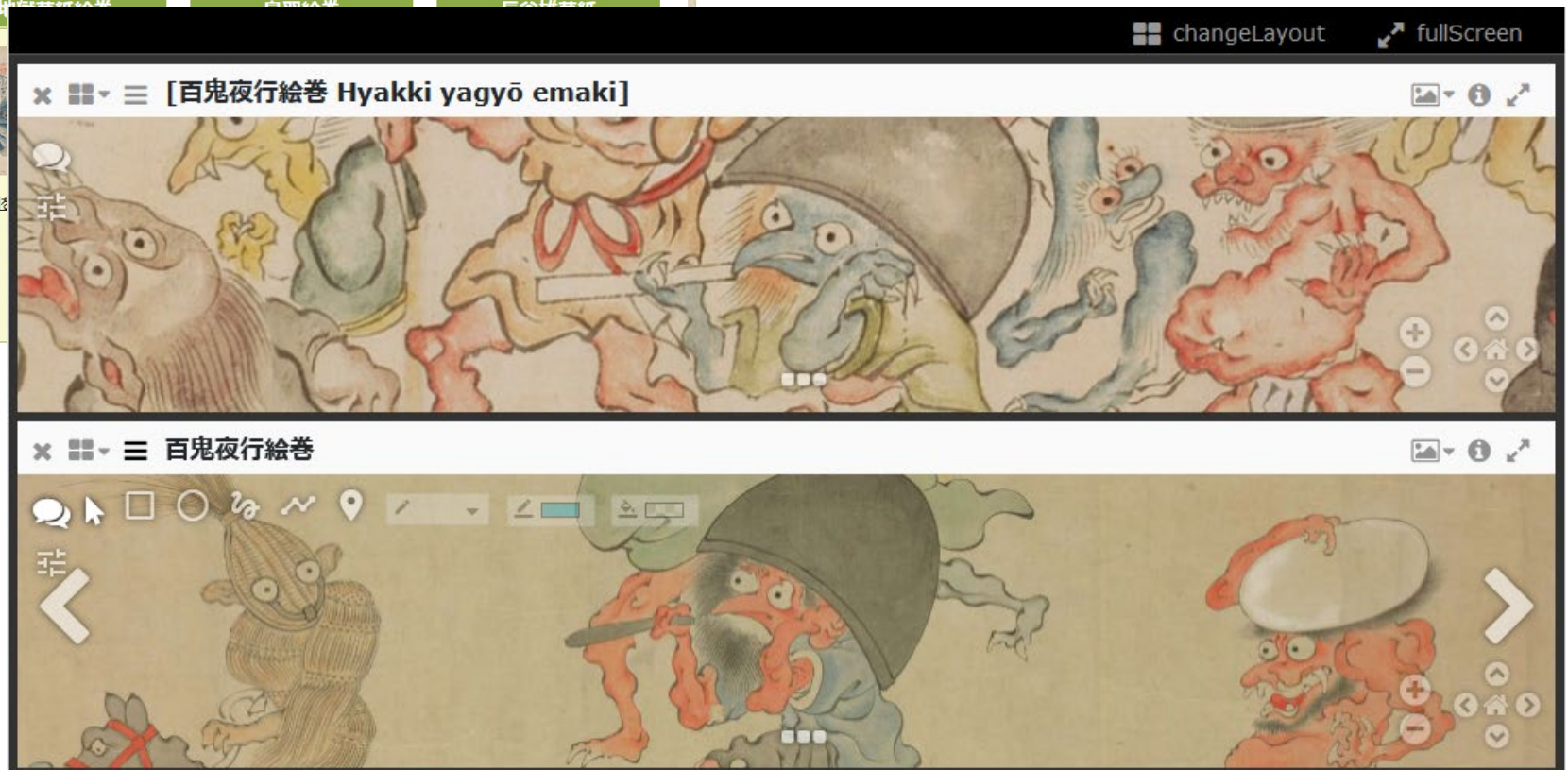


Japan Search (<https://jpsearch.go.jp/>)

画像の相互運用 — IIIF (International Image Interoperability Framework)



- アイコンに画像へのアクセス情報が紐づいている。
- アイコンをビューワにドラッグ & ドロップすることで画像を表示できる。



IIIFの機能を使った画像の比較

特に、情報同士の接点となるようなデータ(空間、時間、人など)

歴史地名データ(人間文化研究機構・H-GIS研究会)

- 242,544件の地名(明治～昭和初期)と経緯度のデータ
- 利用制限なし(出典の明示のみ)
- データのみ公開(検索機能は一切なし)

"ID","地名","属性","形状","緯度","経度","緯度2","経度2","出典","出典詳細","備考"
10000001,"阿位郷",6,1,35.137222,132.944444,,,"大日本地名辞書","3巻 400頁",""
10000002,"阿為神社",12,1,34.854166,135.562777,,,"大日本地名辞書","2巻 573頁 (藍原)",""
10000003,"阿井",5,1,35.137222,132.944444,,,"大日本地名辞書","3巻 400頁 (阿位郷)",""
10000004,"鮎川",22,2,36.153055,138.835555,36.269722,139.046944,"大日本地名辞書","6巻 742頁",""
10000005,"安威郷",6,1,34.853333,135.565000,,,"大日本地名辞書","2巻 573頁",""
10000006,"安威川",22,2,34.981666,135.563055,34.753888,135.528333,"大日本地名辞書","2巻 573頁 (安威郷)",""
10000007,"相川",22,2,35.393333,136.464166,35.324722,136.586944,"大日本地名辞書","5巻 403頁",""

■
■
■

www.hutime.org

www.hutime.org

www.hutime.org

www.hutime.org

技術や規格の標準化の例

TEI (Text Encoding Initiative)

テキストデータの構造化。

CIDOC CRM (International Committee for Document - Conceptual Reference Model)

国際博物館会議 (ICOM) のドキュメンテーション委員会 (CIDOC) が策定
作品と著者などの関係モデルを標準化

Linked Data

セマンティックWeb技術の活用

データの活用に対する意識の浸透

- Open Data
- CCライセンス
- FAIR原則

人文情報学の研究者だけでなく
人文学プロパーの研究者にも

当座の方向

人文学でのデータ駆動型の研究

人文学の情報を活用する(人文学以外も含めて)

課題

- データを作ること、発信することに対する評価
- 人文学の情報を使えるデータサイエンティストの育成
- 異分野間のコミュニケーションの場