



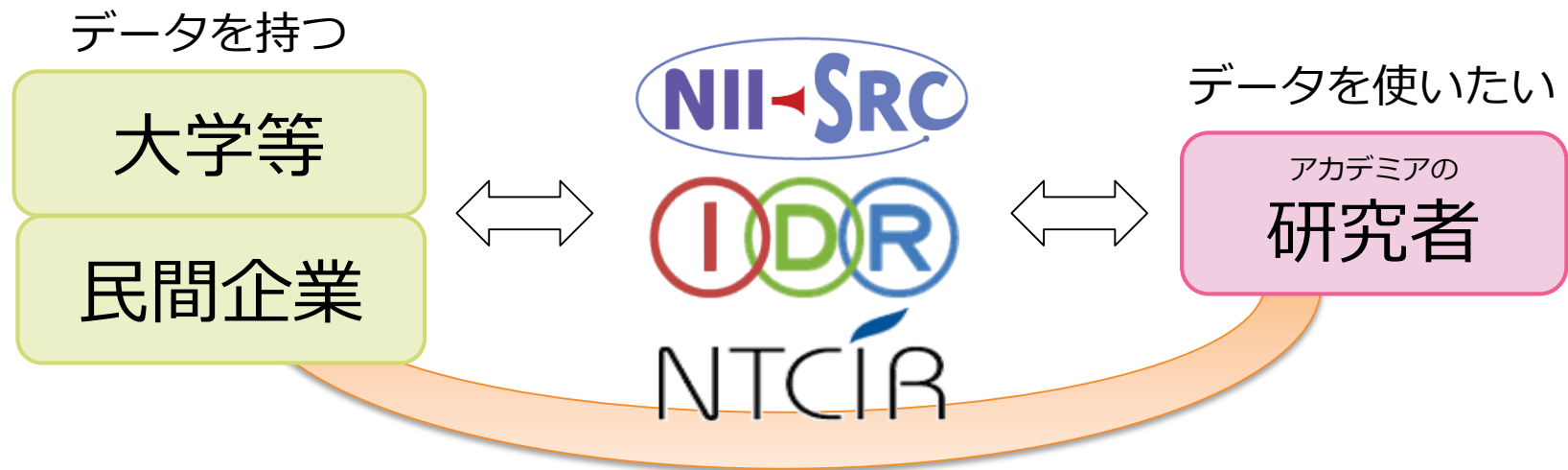
NII-IDRユーザフォーラム2021 ：IDRの活動報告

大須賀智子，大山敬三（IDR事務局）

国立情報学研究所

データセット共同利用研究開発センター

データセット共同利用研究開発センターの主な活動



産学界のデータを学術研究目的で大学等に提供

オープンデータにできないリアルなデータ（個人情報，著作権，…）を，適正な管理の下で，契約に基づき利用可能に

⇒ コミュニティを繋いで研究を活性化



IDRで提供中の民間企業のデータセット

- Yahoo! データセット
- 楽天データセット
- ニコニコデータセット
- リクルートデータセット
- クックパッドデータセット
- LIFULL HOME'Sデータセット
- 不満調査データセット
- Sansanデータセット
- インテージデータセット
- オリコンデータセット
- ダイエットロコミデータセット
- 弁護士ドットコムデータセット
- アットホームデータセット ※2020/12 新規提供開始！

情報学研究データリポジトリ (IDR)

情報学研究データリポジトリ (Informatics Research Data Repository : IDR) は、国立情報学研究所 (NII) のデータセット共同利用研究開発センター (DSC) が運営するデータセットの共同利用事業です。IDRでは各種のデータセットを民間企業や大学等研究者から受け入れて研究者に提供するためのサービスを行っています。

データセットの利用を希望される方は、「[データ一覧](#)」から各データセットのページにアクセスし、記載内容をよくご確認ください。データセットの利用を申請された場合、「[IDRデータセット提供サービス規約 \(法人利用者用\)](#)」(データセットによっては「[個人利用者用](#)」) に同意いただいたものと見なします。

データセットの提供を希望される大学等の研究者の方は「[大学等研究者提供データセット受入](#)」のページをご覧ください。また、民間企業等の場合は、「[問い合わせ](#)」に記載のIDR事務局にご相談ください。

音声コーパス	2021-03-12	「大阪大学 マルチモーダル対話コーパス」を追加しました。→ NIIトピックス
研究者提供データ	2021-02-10	「楽天データセット」に新規データを追加しました。→ NIIトピックス
グループコミュニケーション ・ ションコーパス (TDO-)	2020-12-15	「アットホームデータセット」の提供を開始しました。→ プレスリリース

→ 展示会場企業ブースにポスター
お昼にはライトニングトークも！

例：弁護士ドットコムデータセット

※2020年3月
提供開始！

オンライン法律相談サービス「みんなの法律相談」に投稿された約25万件の質問とそれに対する弁護士の回答のテキストデータ

- 質問のカテゴリ
- 「弁護士が同意」
の人数
- 質問者が付与した
「ありがとう」
「ベストアンサー」



<https://www.bengo4.com/> のトップページより引用

例：アットホームデータセット

※2020年12月提供開始！

アットホーム社運営の全国不動産情報ネットワーク（加盟・利用不動産
店数58,000店以上 ※2020.12時点）に登録された全国の不動産物件情報



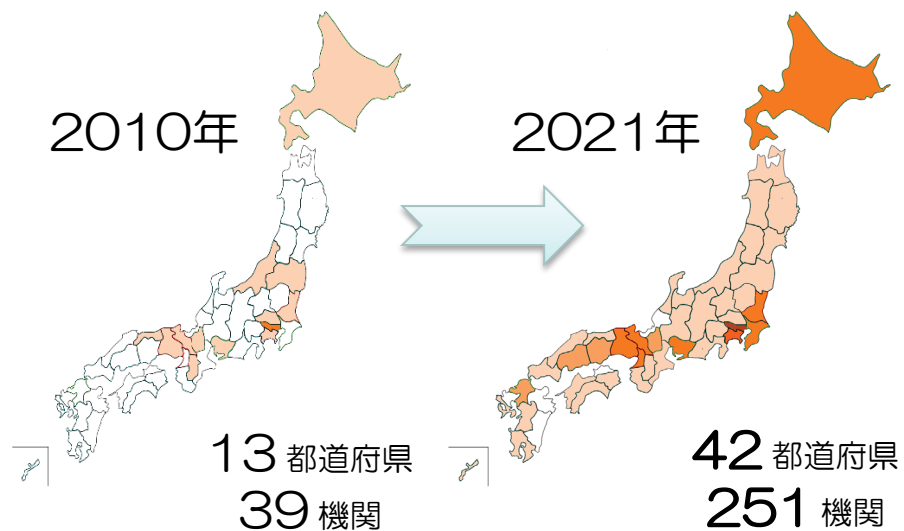
- ・ 2015年～5年間の月次データ
- ・ 種別
 - ― 賃貸居住用，賃貸事業用
 - ― 売買居住用
- ・ 項目
 - ― 賃料または価格
 - ― 間取り
 - ― 建物構造，築年
 - ― 立地
 - ― 最寄沿線・駅，徒歩分
 - ― 諸設備

<https://www.athome.co.jp/> のトップページより引用

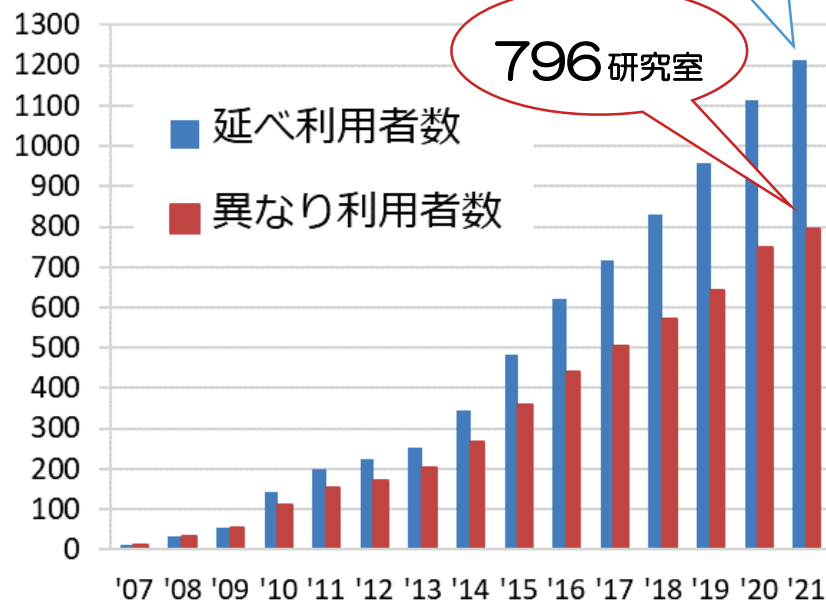
企業データ提供実績：研究室単位（～2021.9）

データ提供先

（ニコニコ，Sansanデータセット除く）



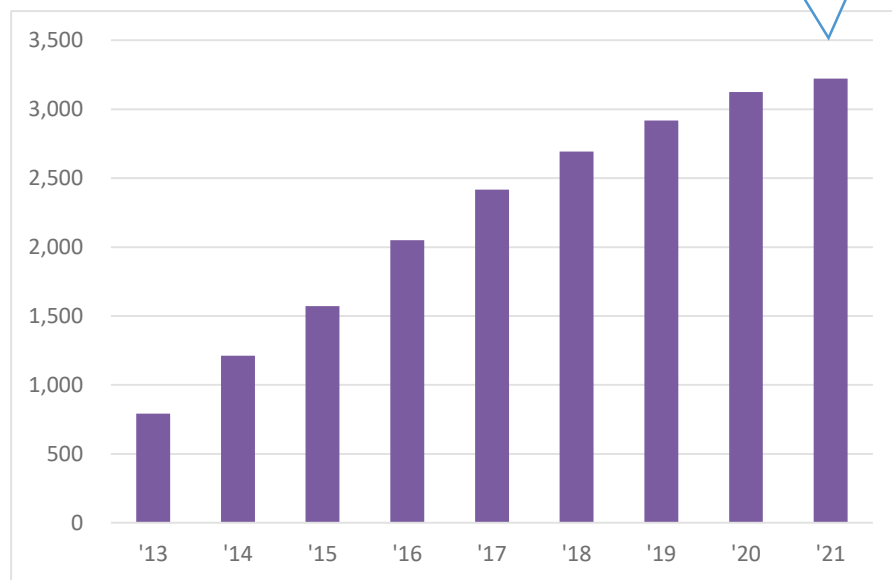
利用者数の推移



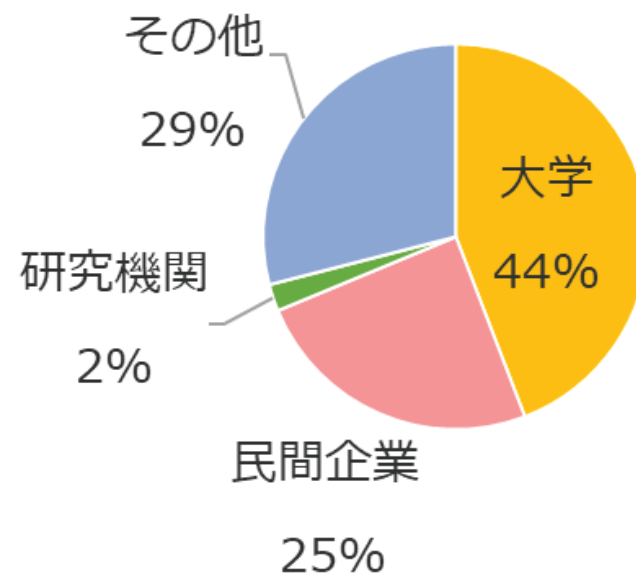
企業データ提供実績：個人単位（～2021.9）

利用申請者数の推移

- ニコニコデータセット
- Sansanデータセット
- カテゴリ別不満特徴語辞書



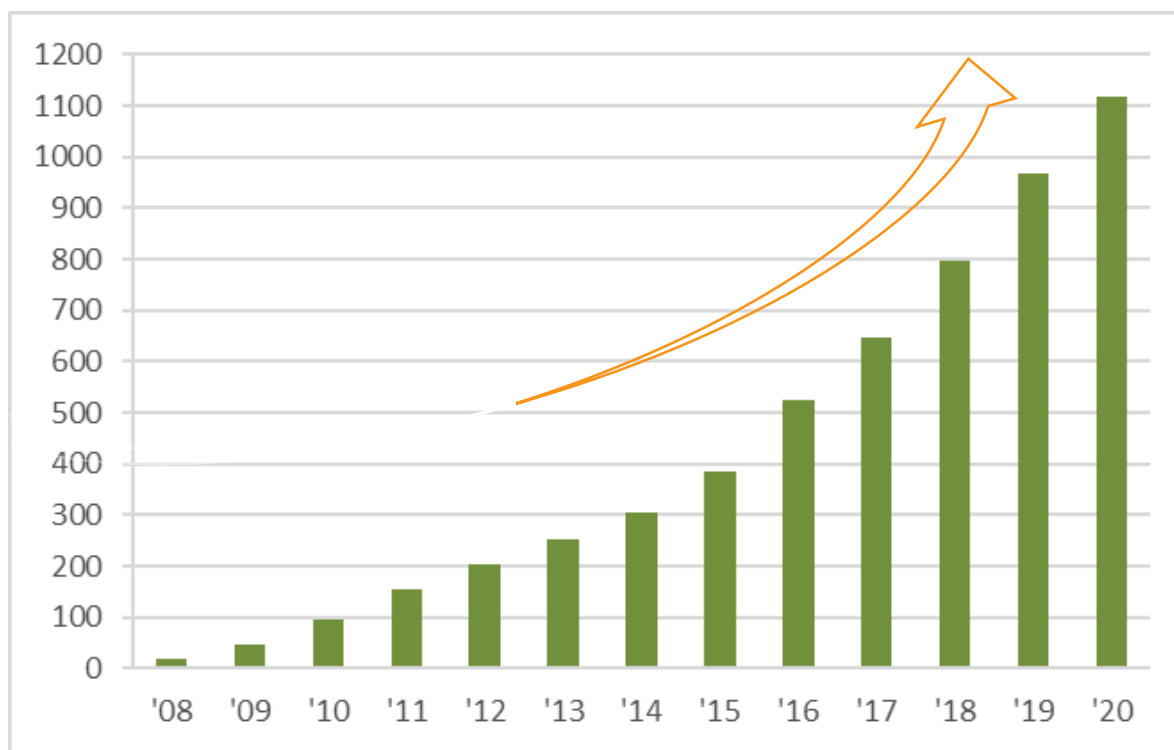
利用者の所属



企業データを利用した研究発表（～2020年度）

研究成果外部発表数の推移

※毎年度の利用報告ベース



データを利用した論文リストは公開しています

研究発表

研究発表

<https://dsc.repo.nii.ac.jp/>



データセットのデータDOIを取得！

ログイン English

DSC Reference Portal

DSCホーム IDRホーム SRCホーム NTCIRホーム

データセット共同利用研究開発センター（DSC）から提供するデータセットに関する文献情報のポータルです。
以下の情報を公開しています。（SRCの音声コーパス及びNTCIRテストコーパス）

- 利用者から報告があった研究成果やデータセットを紹介した参考文献
- DSC提供のデータセットのメタデータ、DOI、謝辞、引用方法

トップ ランキング

全文検索

インデックスを選択してください

インデックスツリー

- 研究発表
- データセット
 - IDRデータセット
 - Yahoo!データセット
 - 楽天データセット
 - ニコニコデータセット
 - リクルートデータセット
 - クックパッドデータセット
 - LIFULL HOME'Sデータセット
 - 不満調査データセット
 - Sansanデータセット
 - インテージデータセット
 - オリコンデータセット
 - ダイエツロコミデータセット

研究者提供データ

お問合わせ

国立情報学研究所 データセットリポジトリ
Copyright(C)2010-2011

(※試験公開中)

Permalink : <http://doi.org/10.32130/idr.10.1>

IDR 顧客満足度調査データ

JaLC DOI	info:doi/10.32130/idr.10.1
アイテムタイプ	IDR dataset
言語	日本語
サブジェクト	アンケート
提供者	株式会社oricon
配布者	国立情報学研究所 データセットリポジトリ
提供日	
収録期間	
データ形式	
概要	
収録データセット	
利用資格	
申請手続	利用申請書を提出し、審査を経て同意書を提出。
利用許諾	利用規約及びサービス規約による。
引用記載例	株式会社oricon ME (2019): 顧客満足度調査データ。国立情報学研究所情報学研究データリポジトリ。(データセット)。 https://doi.org/10.32130/idr.10.1
引用記載例(英)	oricon ME Inc. (2019): Customer satisfaction survey data. Informatics Research Data Repository, National Institute of informatics. (dataset). https://doi.org/10.32130/idr.10.1

データセット

データDOIの例：
オリコン顧客満足度調査データ
<http://doi.org/10.32130/idr.10.1>

データ利用者の方へ：研究発表の際は…

【投稿する前に】

- 謝辞への記載（※）
- 参考文献へのデータDOI
- 原稿の事前提出（データ）

【毎年度末に】

- 報告書，論文のリスト

にご協力お願いします！

IDR 顧客満足度調査データ	
    	
JaLC DOI	info:doi/10.32130/idr.10.1
アイテムタイプ	IDR dataset
言語	日本語
サブジェクト	アンケート調査データ
提供者	株式会社oricon ME
配布者	国立情報学研究所 情報学研究データリポジトリ
提供日	2019-07-01
収録期間	2016 - 2019
データ形式	text(csv), Excel(xlsx)
概要	株式会社oricon MEが実施している「オリコン顧客満足度」の調査データ。「保険」「住宅」など11ジャンル、88業種について、2016年以降に実施された200件のアンケート調査の個票を含む。調査項目は、利用経験、満足度、サービスを利用する際の重視点、他者への推奨意向、再利用意向などのほか、対象となる業種ごとに異なる。自由記述の回答は削除済み。回答者数は延べ67万名で、性別や年齢など回答者の属性も付随。
収録データセット	オリコンデータセット
利用資格	大学及び公的研究機関の研究者。その他利用条件あり。
申請手続	利用申請書を提出し、審査を経て同意書を提出。
利用許諾	利用規約及びサービス規約による。
引用記載例	株式会社oricon ME (2019): 顧客満足度調査データ. 国立情報学研究所情報学研究データリポジトリ. (データセット). https://doi.org/10.32130/idr.10.1
引用記載例(英)	oricon ME Inc. (2019): Customer satisfaction survey data. Informatics Research Data Repository, National Institute of Informatics. (dataset). https://doi.org/10.32130/idr.10.1
謝辞記載例	本研究では、国立情報学研究所のIDRデータセット提供サービスにより株式会社oricon MEから提供を受けた「オリコンデータセット」を利用した。
謝辞記載例(英)	In this paper, we used "Oricon Dataset" provided by oricon ME Inc. via IDR Dataset Service of National Institute of Informatics.
提供情報	NII IDR: オリコンデータセット
提供状況	提供中

※ 記載方法 ⇨ <https://dsc.repo.nii.ac.jp/>

CiNii Research

CiNii Research

フリーワード

検索

▼ 詳細検索

クックパッドデータ

 DOI

 Web Site

 <https://info.cookpad.com/>
作成者

メタデータ

| 作成日 2015-02-14
| 著作権 クックパッド株式会社が定めた利用規約に基づき作成されたデータです。Based on the terms of use st
| 公開日 9999
| 公開者 国立情報学研究所 情報学研究フ

関連論文

レシピサイトからの調理のコツ抽出手法の提案

米田,吉希,灘本,明代 DEIM2018 2018

 Web Site

 研究データあり

 参考文献1件

使用食材に注目した複数レシピの路線図化

紺屋,夏月,丸山,一貴 情報処理学会 インタラクシオン2018 517-522, 2018

 Web Site

 研究データあり

 参考文献1件

組み合わせを考慮した献立検索手法の提案

南谷,悠大,前田,亮 第10回データ工学と情報マネジメントに関するフォーラム (DEIM2018) 論文集 2018

 Web Site

 研究データあり

 参考文献1件

Mining online activity data to understand food consumption behavior: a case of Asian fish sauce among Japanese consumer

Mitsutoshi,Nakano,Hiroaki,Sato,Toshihiro,Watanabe,Katsumi,Takano,Yoshimasa,Sagane Food Science and Nutrition 2018

 DOI

 Web Site

 研究データあり

 参考文献1件



IDRで提供中のその他のデータセット

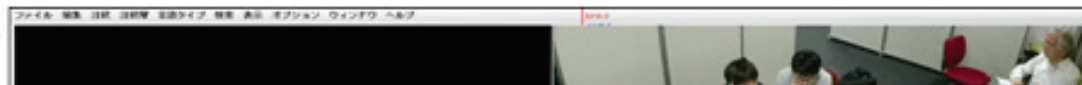
- NTCIRテストコレクション
 - NTCIRプロジェクトが構築
 - タスクデータ36種類（課題と正解），文書データ2種類
- 音声コーパス
 - 音声資源コンソーシアムが受入
 - 音声のデータセット46種類
- 研究者等提供データセット

本日3件発表あり！
(P17, P25, S10)

2019年～
IDRの新たな取組

研究者等提供データセットの例（1）

- ・グループコミュニケーションコーパス ※2019年8月提供開始



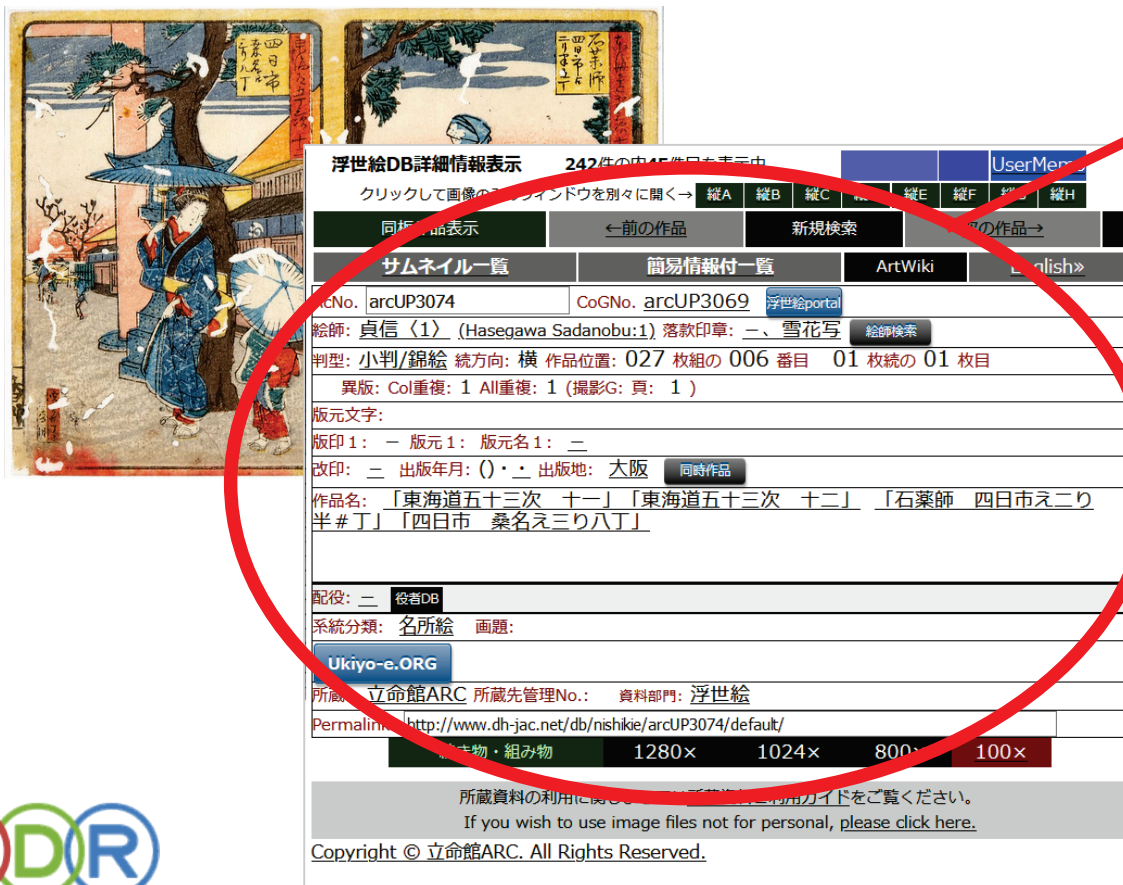
この他に

- ・システムとの対話の映像データ
「大阪大学マルチモーダル対話コーパス」
※2020年8月提供開始
- ・手話の映像データ
「工学院大学 多用途型日本手話言語DB」
※2021年6月提供開始（NEW）



研究者等提供データセットの例（2）

- 立命館ARC所蔵浮世絵データベース ※2020年3月提供開始



浮世絵DB詳細情報表示 242件の中15件を表示中

クリックして画像のウィンドウを別々に開く → 縦A 縦B 縦C 縦D 縦E 縦F 縦G 縦H

同作品表示 ← 前の作品 新規検索 後の作品 →

サムネイル一覧 簡易情報付一覧 ArtWiki English»

arcNo. arcUP3074 CoGNo. arcUP3069 浮世絵portal

絵師: 貞信 〈1〉 (Hasegawa Sadanobu:1) 落款印章: 一、雪花写 絵師検索

判型: 小判/錦絵 続方向: 横 作品位置: 027 枚組の 006 番目 01 枚続の 01 枚目

異版: Col重複: 1 All重複: 1 (撮影G: 頁: 1)

版元文字:

版元 1: - 版元 1: 版元名 1: 二

改印: 二 出版年月: ()・二 出版地: 太阪 同時作品

作品名: 「東海道五十三次 十一」 「東海道五十三次 十二」 「石薬師 四日市え二り半 # 丁」 「四日市 桑名え三り八丁」

配役: 二 役者DB

系統分類: 名所絵 画題:

Ukiyo-e.ORG

所蔵: 立命館ARC 所蔵先管理No.: 資料部門: 浮世絵

Permalink: <http://www.dh-jac.net/db/nishkie/arcUP3074/default/>

主物・組み物 1280× 1024× 800× 100×

所蔵資料の利用に際しては、必ず利用ガイドをご覧ください。
If you wish to use image files not for personal, please click here.

Copyright © 立命館ARC. All Rights Reserved.

浮世絵の書誌情報
約11,000件
(画像データURL付き)

本日発表あり！
(P24)

立命館大学アート・リサーチセンターの
サイトより引用

研究者等提供データセットの例（3）

- 理研記述問題採点データセット ※2020年7月提供開始

記述答案採点の自動化に関する研究

[Top Page](#)[Projects](#)[Resources](#)

プロジェクト概要

記述式問題を通した国語学習は、論理的思考力・判断力・表現力を養うために効
の学習の過程では添削者からのフィードバックが必要ですが、記述答案の添削が
ているという問題があります。我々のチームでは、記述式問題を通した国語学習
び添削の負担軽減を目的とし、記述答案を自動で採点する技術の研究・開発を行
業から提供された採点データを基に自動採点のためのデータセットを整備し、公
発展に貢献することを目指しています。

デモビデオ



研究者等提供データセットの受入手続き

- 受入要項を公開
- 流れ：

受入依頼書の提出



NIIで審査・手続き



提供同意書を提出

NII 国立情報学研究所 National Institute of Informatics

情報学研究データリポジトリ

HOME データ一覧 研究成果一覧 ユーザフォーラム 組織 関連リンク 問い合わせ

> HOME > IDRサービス

大学等研究者提供データセット受入

2019/09/26 作成

情報学研究データリポジトリ（IDR）では、大学等の研究者（作成者）が研究用に作成した各種データセットを他の研究者（利用者）へ提供するための支援サービスを行っています。

本サービスでは、著作権や個人情報保護など様々な理由によりオープンデータとしての公開が難しいデータセットでも、作成者と協力して利用者の範囲、利用目的や利用方法などの条件を定め、IDRを通じて利用許諾契約を締結した利用者にデータセットを提供することによって、できるだけ作成者の負担を軽減しつつ、多くの研究者がデータセットを利用できるようにします。

受入条件

学術研究の対象として高い価値を有し、相当数の利用者が見込まれるデータセットが、本サービスの受入の対象となります。本サービスはアーカイブではありませんので、保存を目的とした受入はしません。

法令上または研究倫理上の観点から、IDRを通じた提供に必要な条件を満たさない場合や、IDRの能力を超える設備や取扱いが要求される場合は、受入をできません。

詳細については下記「書類一覧」の「研究者等提供データセット受入要項」をご覧ください。また、不明な点

3条の「提供

人によって適

の場合、その

またがる場合な

とは、予めご相談ください。

2. 「研究者等提供データセット受入要項」の内容に同意の上、下記「書類一覧」の「データセット受入依

頼書」に必要事項を記入し、添付資料とともに、電子メールの添付ファイルで下記「問い合わせ窓口

民間企業様の場合は個別対応としていますので、
お気軽にメールでお問い合わせください！



ぜひIDRをご利用ください！

- データセットをご利用になりたい方
- データセットの提供をお考えの研究者の方

サイト：<https://www.nii.ac.jp/dsc/idr/>



- データセットの提供をお考えの企業の方

メール：idr@nii.ac.jp

本日の「IDRユーザフォーラム」は…



データの提供者と利用者が一堂に会する場

- ポスターセッション
- 企業セッション
- 招待講演
- スタートアップセッション
- 口頭発表セッション

昨年に引き続き
日本DB学会
データチャレンジ
と一体開催！