



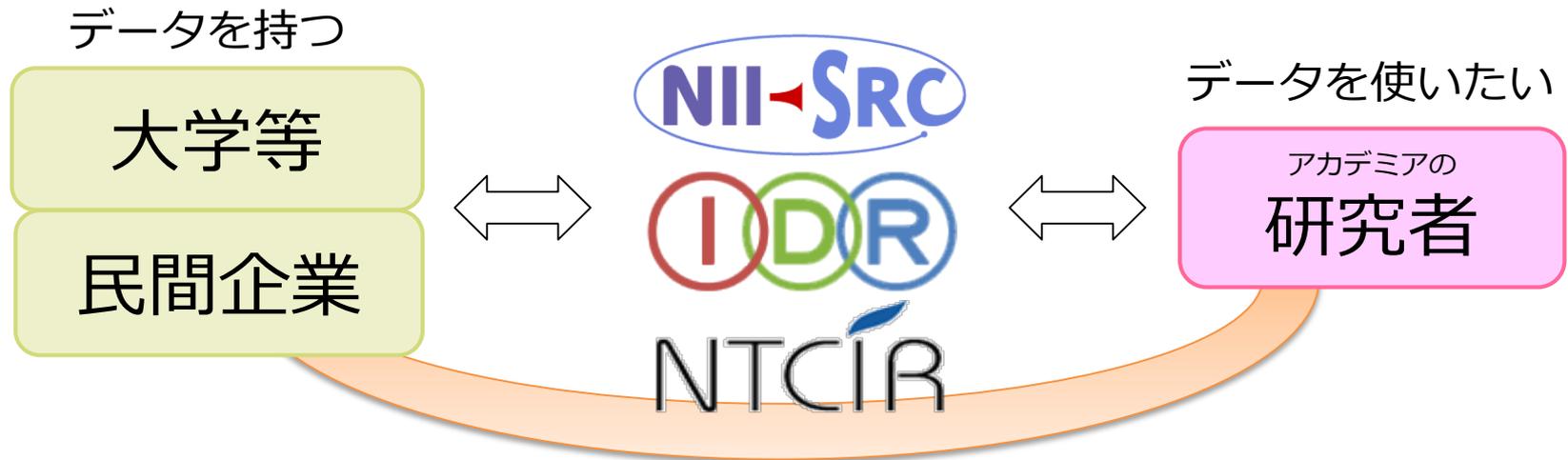
# NII-IDRユーザフォーラム2022 ：IDRの活動報告

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
- ダイエット口コミデータセット
- 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
- JASTメディカルデータセット
- トリガーデータセット
- みんなの評判口コミデータセット ※2022/8 新規提供開始！
- メルカリデータセット ※2022/9 新規提供開始！
- 地球の歩き方旅行記データセット ※2022/11 新規提供開始！

The screenshot shows the homepage of the National Institute of Informatics (NII) Information Research Data Repository (IDR). The header includes the NII logo and the text '情報学研究所 国立情報学研究所 National Institute of Informatics'. The main navigation bar contains links for HOME, データ一覧 (Data List), 研究成果一覧 (Research Results List), ユーザーフォーラム (User Forum), 組織 (Organization), 関連リンク (Related Links), and お問い合わせ (Contact Us). A red box highlights the '民間企業提供データ' (Data provided by private companies) section, which lists various data sets including Yahoo!, Rakuten, Nico Nico, Recruit, Cookpad, LIFULL HOME'S, and Sansan. Another red box highlights the '更新情報' (Update Information) section, which lists recent updates such as 'オリコンデータセット' (Oricon Data Set) updated on 2022-06-07, '音声コーパス『日本語共感的音声対話コーパス (STUDIES)』' (Audio Corpus) provided on 2022-03-29, 'JASTメディカルデータセット' (JAST Medical Data Set) updated on 2022-03-25, and 'トリガーデータセット' (Trigger Data Set) provided on 2022-03-15.

→ 展示会場に各企業のポスターあり  
お昼にはプレゼンテーション・  
インタラクティブ発表も！

# 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」

The screenshot shows the homepage of lawyer.com. At the top, there is a navigation bar with the logo and user options like '閲覧履歴' (View History), 'お気に入り' (Favorites), and 'メニュー' (Menu). Below the header, a yellow banner contains the text: 「弁護士ドットコム」は日本最大級の法律相談ポータルサイトです (Lawyer.com is the largest legal consultation portal site in Japan). To the right of this banner is a button for 「弁護士専用ログイン」 (Lawyer-only login). Below the banner, there are two main search sections. The first is for finding lawyers, with a search bar and a button labeled 「弁護士を探す」 (Find lawyer). The second is for legal Q&A, with a search bar and a button labeled 「法律Q&Aを探す」 (Find legal Q&A). To the right of these search sections, there are statistics: 「会員弁護士数 19,567人」 (19,567 member lawyers) and 「相談数 1,083,162件」 (1,083,162 consultations). Below the search sections, there is a section for 「お悩みから解決方法を探す」 (Find solutions from your concerns) with icons for 「離婚・男女問題」 (Divorce/Gender issues), 「借金」 (Debt), and 「相続」 (Inheritance). On the right side, there is a section for 「弁護士ドットコムニュース」 (Lawyer.com News) with a list of recent articles.

<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

The screenshot shows the homepage of the 'at home' real estate portal. At the top, it displays the company logo and key statistics: '現在 2,381,958 件 公開中!' (Currently 2,381,958 items published!) and 'アットホーム加盟・利用不動産店数 59,543 店 2021年11月1日現在' (At Home franchise/used real estate store count 59,543 stores as of 2021.11.01). Navigation icons for 'お気に入り' (Favorites), 'お気に入り' (Favorites), and 'お気に入り' (Favorites) are visible, along with a 'ログイン' (Login) button. The main content area is divided into 'RENT 借りる (賃貸)' (Rent/Rental) and 'BUY 買う' (Buy) sections. The 'RENT' section lists options like '賃貸' (Rental), '賃貸店舗' (Rental shop), '賃貸土地' (Rental land), and '賃貸ビル・倉庫・その他' (Rental building, warehouse, etc.). The 'BUY' section lists '買うをまとめて探す' (Search for things to buy together), 'マンション' (Mansion), '一戸建て' (Single house), and '売土地' (Land for sale). A 'アプリで探す' (Search with app) section promotes the mobile app. At the bottom, there are icons for '売る' (Sell), '不動産会社を探す' (Find real estate company), '引越す' (Move), and '調べる' (Check/Research). A cartoon illustration of a man and a woman in red and white striped shirts is at the bottom right.

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

- 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
  - など

# 例：JASTメディカルデータセット

※2022年1月  
提供開始！

レセプトデータから、傷病情報×性別・年代・都道府県別に  
患者数・レセプト件数・医療費を月次単位で集計したデータ

## JASTメディカルデータセット項目

項目名称 (和名)	変数名称	項目説明
診療年月	medtreat_month	201904, 201905, 201906
ICD-10 章分類	icd10_dai	
ICD-10 中間分類	icd10_chu	ICD
ICD-10 3桁分類	icd10_sho	
性別	sex_type	01_
年代	age_kbn	01_
医療機関都道府県	pref_type	01_
レセプト件数	rezept_count	当該
患者数	patient_count	当該
主傷病寄せ医療費	syu_medical_cost	当該
主傷病フラグ	syu_flag	主傷

## JASTメディカルデータセット解析例

### コロナ禍における呼吸器系疾患患者の実態について

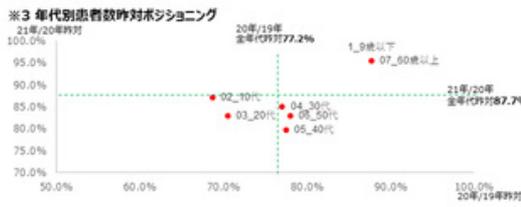


※2 2019年～21年 ICD-10別 1～6月(6ヵ月)の患者数対比

ICD-10呼吸器系分類	20年/19年	21年/20年
呼吸器全般	77.2%	87.7%
急性鼻咽喉炎	67.3%	78.1%
血管運動性鼻炎及びアレルギー性鼻炎	80.1%	99.7%
喘息	78.9%	89.2%

※1 呼吸器系疾患の患者数は、コロナの発生に合わせて減少する傾向にありますが、その後もコロナのピーク時には下がる傾向となっております。これは受診控えの傾向が出現している可能性があります。

※2 ICD-10呼吸器系の分類別で患者数の対比を見ても減少傾向は明らかです。急性だけではなく慢性も同様の傾向となっております。重症化するとより治療費が必要となってくる喘息なども傾向が一緒の為、注視していかなければなりません。



※3 年代別の対比で減少幅に特徴がないかを見てみると、特に10代～20代が顕著に見えます。逆に小児世代やシニア世代は対比は落ちますが、減少幅としては小さいものとなります。



# 例：トリガーデータセット

※2022年3月提供開始！

文化庁若手アニメーター育成プロジェクト「アニメミライ2013」での  
劇場公開アニメ『リトルウィッチアカデミア』制作時の中間素材



- シナリオ・台本
- 絵コンテ
- 美術
- 設定
- 色彩
- カット袋・タイムシート  
・レイアウト・原画
- 仕上げ

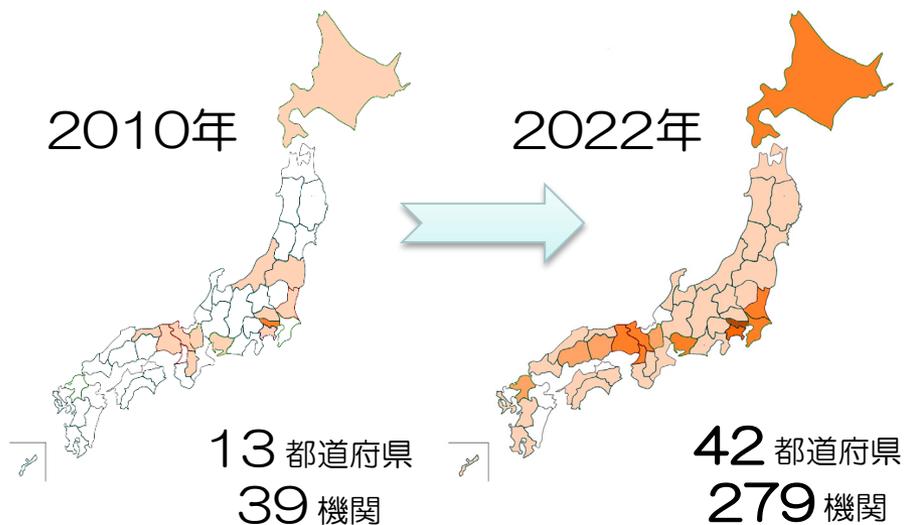
(動画素材, 映像は含まれません)

©TRIGGER Inc.

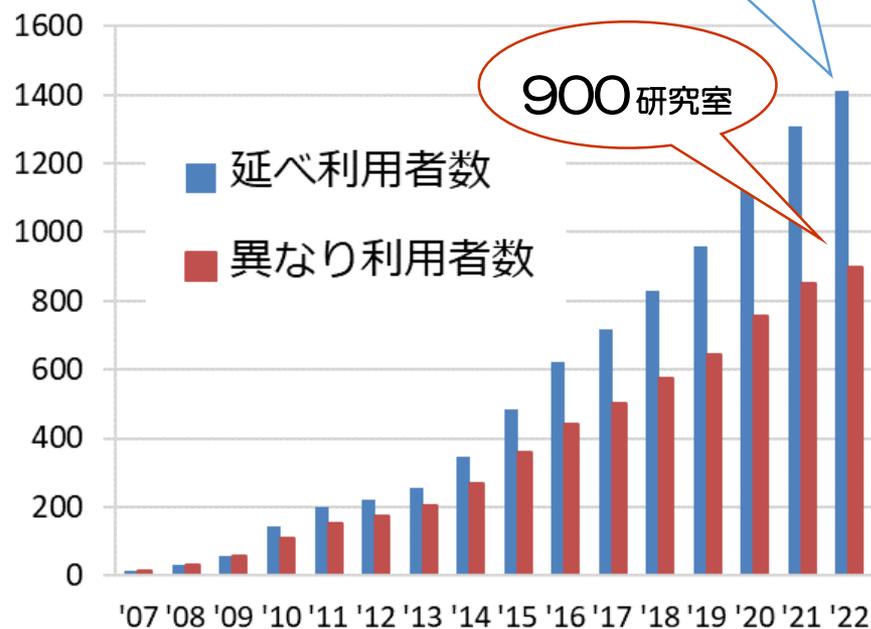
# 企業データ提供実績：研究室単位（～2022.9）

## データ提供先

（ニコニコ， Sansanデータセット除く）



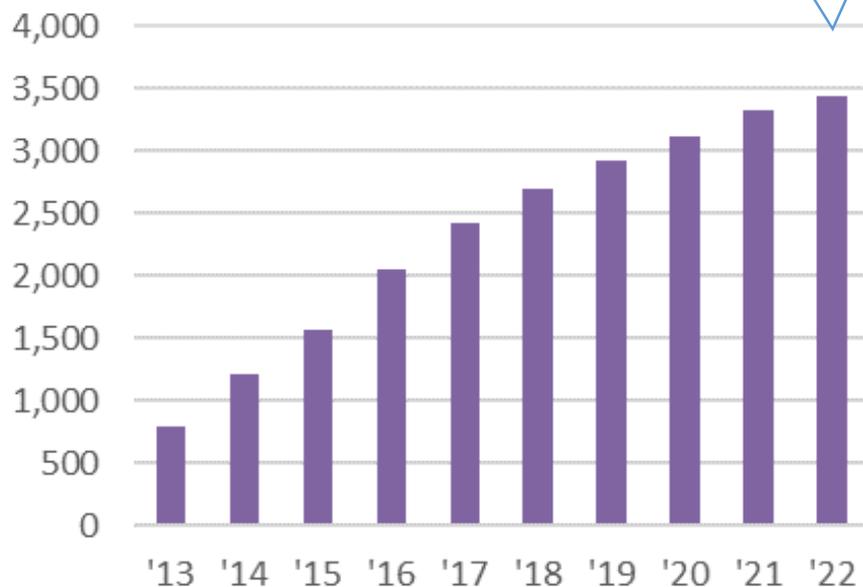
## 利用者数の推移



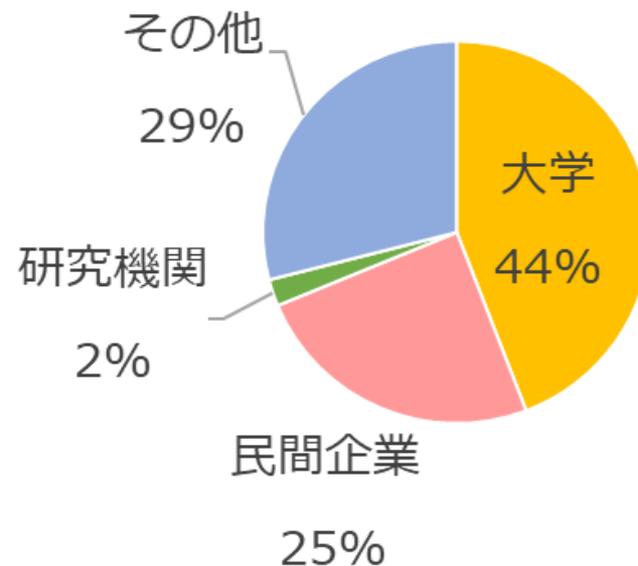
# 企業データ提供実績：個人単位（～2022.9）

## 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



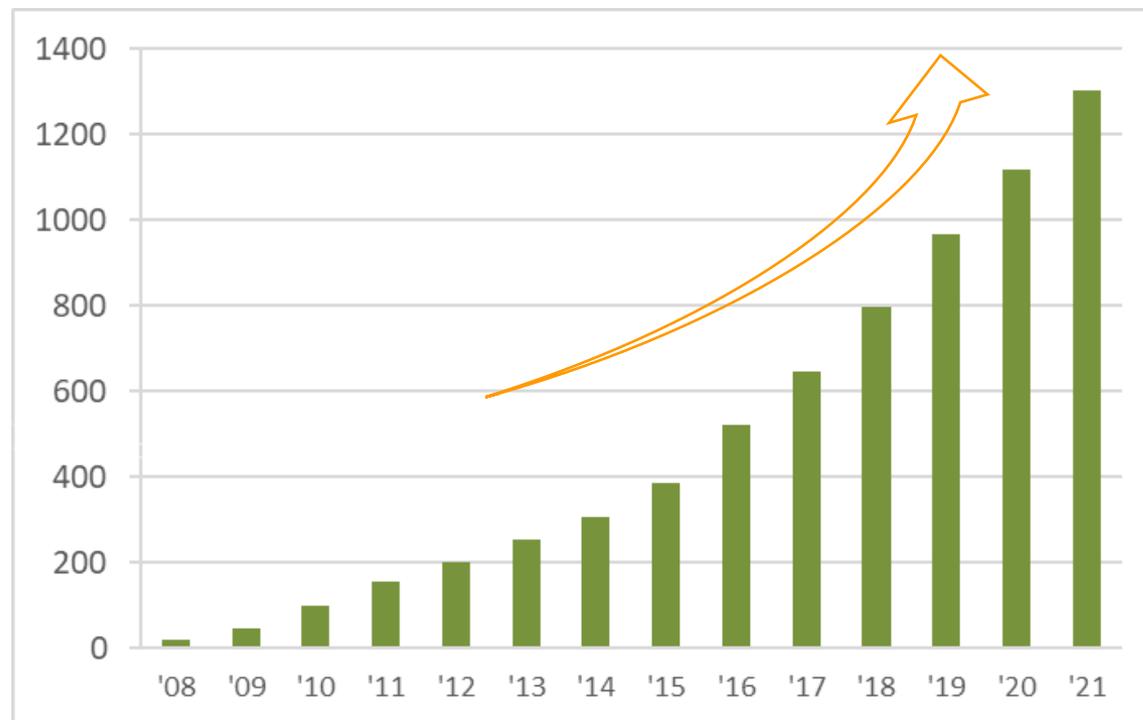
## 利用者の所属



# 企業データを利用した研究発表（～2021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

情報学研究データリポジトリ

研究発表

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



# データセットのデータDOIを取得！

The screenshot shows the DSC Reference Portal interface. On the left, the 'インデックスツリー' (Index Tree) is expanded to 'データセット' (Dataset), and 'オリコンデータセット' (Oricon Dataset) is highlighted. A red box highlights this item, with an arrow pointing to the right. On the right, the '顧客満足度調査データ' (Customer Satisfaction Survey Data) page is shown. A red box highlights the 'JaLC DOI' field, which contains the value 'info:doi/10.32130/idr.10.1'. A large red callout box on the right contains the text: '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'. The URL 'http://doi.org/10.32130/idr.10.1' is also visible in the top right corner of the page.



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

# 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  
にご協力をお願いします！

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	<a href="#">株式会社oricon ME</a>
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). <a href="https://doi.org/10.32130/idr.10.1">https://doi.org/10.32130/idr.10.1</a>
引用記載例(英)	oricon ME Inc. (2019): Customer satisfaction survey data. Informatics Research Data Repository, National Institute of Informatics. (dataset). <a href="https://doi.org/10.32130/idr.10.1">https://doi.org/10.32130/idr.10.1</a>
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	<a href="#">NII IDR: オリコンデータセット</a>
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 Based on the terms of use st
公開日	9999
公開者	国立情報学研究所 情報学研究ラ

## 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48種類

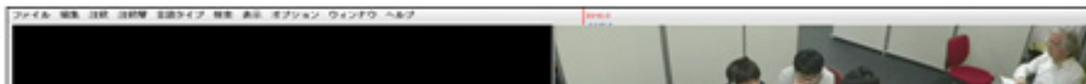
本日3件発表あり！  
(P21, P22, S16)

- 研究者等提供データセット

2019年～  
IDRの新たな取組

# 研究者等提供データセットの例

- 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」

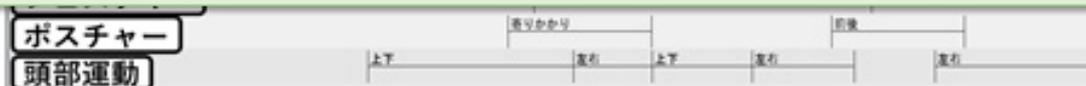
※2020年8月提供開始

- 手話の映像データ

「工学院大学 多用途型日本手話言語DB」

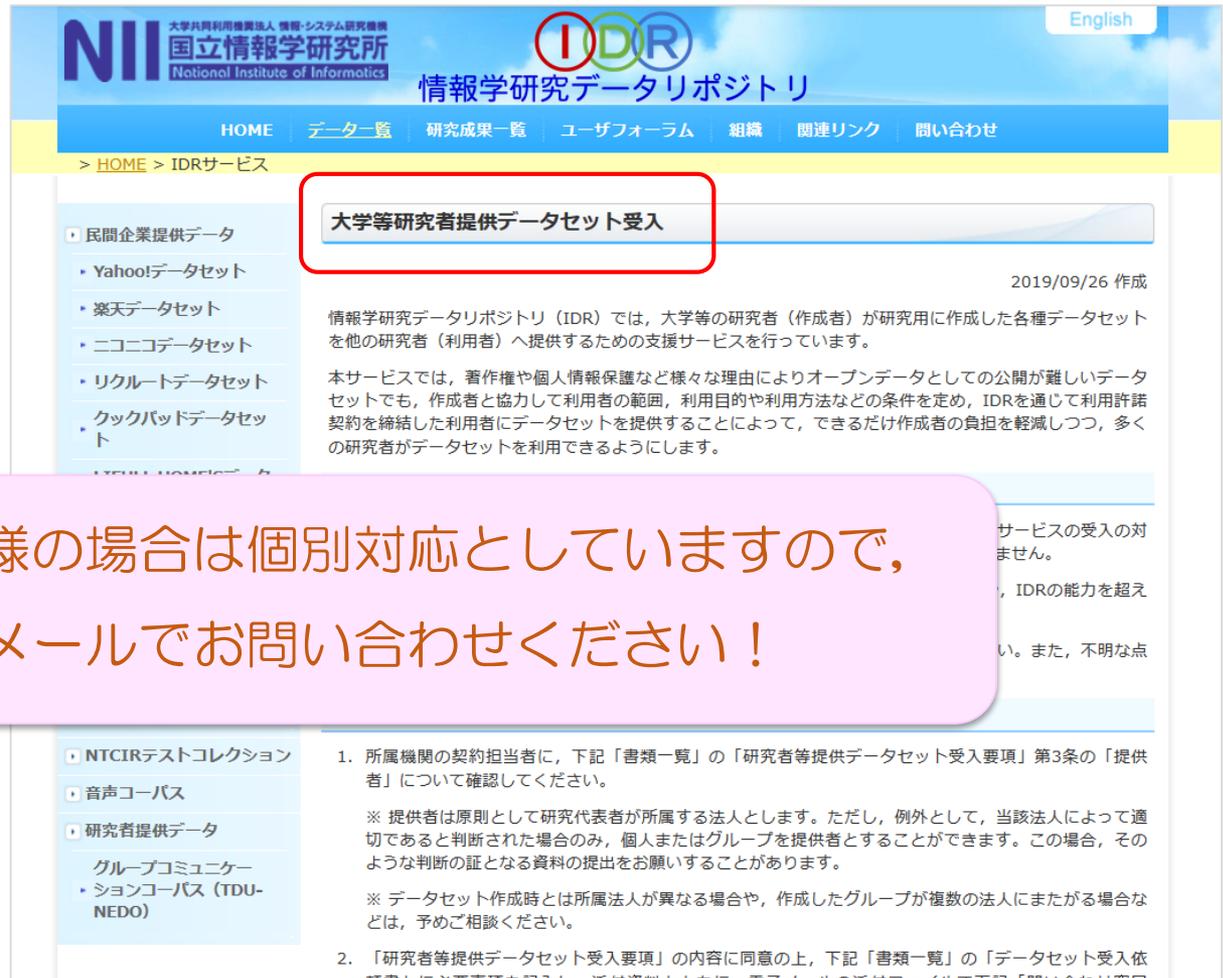
※2021年6月提供開始

本日発表あり！  
(P20)



# 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



The screenshot shows the NII IDR website interface. The header includes the NII logo and the text '国立情報学研究所 National Institute of Informatics'. The IDR logo is prominently displayed. The main navigation bar contains links for HOME, データ一覧, 研究成果一覧, ユーザーフォーラム, 組織, 関連リンク, and お問い合わせ. Below the navigation bar, the breadcrumb trail reads '> HOME > IDRサービス'. The main content area features a red-bordered box around the link '大学等研究者提供データセット受入'. To the left, a sidebar lists various data sets: 民間企業提供データ (including 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), NTCIRテストコレクション, 音声コーパス, and 研究者提供データ (including グループコミュニケーショントラjectoryコーパス (TDU-NEDO)). The main text explains that IDR provides support for researchers to share their data sets. It states that while open data is preferred, IDR facilitates sharing under specific conditions through a license agreement. A list of terms and conditions is provided at the bottom of the page.

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。

1. 所属機関の契約担当者に、下記「書類一覧」の「研究者等提供データセット受入要項」第3条の「提供者」について確認してください。

※ 提供者は原則として研究代表者が所属する法人とします。ただし、例外として、当該法人によって適切であると判断された場合のみ、個人またはグループを提供者とすることができます。この場合、そのような判断の証となる資料の提出をお願いすることがあります。

※ データセット作成時とは所属法人が異なる場合や、作成したグループが複数の法人にまたがる場合などは、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「お問い合わせ窓口」

民間企業様の場合は個別対応としておりますので、  
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](mailto: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！