



NII-IDRユーザフォーラム2019 ：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

IDR発足の背景

データを使いたい
研究者

ビッグデータ, AI ... ⇒ 研究と実応用が接近

【従来】 人手で集めたデータ, 実験室で作ったデータ

【現在】 実サービスや社会で発生するリアルで大量のデータ

データを持つ
産業界

アカデミアへのデータ提供の動機はある

社会貢献, 現状の課題解決, 人材の確保, ...

- 大学と共同研究 → ツテが必要, 個別対応は大変
- 自社ビジネスへのリスク?

ただし, リアルデータには様々なしがらみが
(個人情報, プライバシー, 著作権, 営業秘密...)

⇒ IDR：情報学研究データリポジトリ



情報学研究用データセットの共同利用事業

- 2010年より運用開始
- データセット提供（収集，受入，配布）
：セミクローズドなデータの提供

⇒ 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 ※2019/6 新規提供開始!
- ダイエットロコミデータセット ※2019/10 新規提供開始!

The screenshot shows the homepage of the Information Research Data Repository (IDR) at NII. The page features a navigation menu with 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, NicoNico, Recruit, Cookpad, LIFULL HOME'S, Sansan, Intage, and Oricon. Below this, the 'お知らせ' (Notice) section contains several updates, including the expansion of the IDR User Forum 2019 on November 29, the start of data provision for the Diet Ro-comi dataset on October 29, 2019, and the start of data provision for the Oricon dataset on June 6, 2019.



オリコンデータセット ※6月提供開始！

満足度調査データ

- 200調査, 67万回答
- 11ジャンル・88業種
 - 保険
 - マネー
 - 通信
 - 生活
 - ショッピング
 - トラベル
 - 住宅
 - 受験・スクール
 - 転職
 - 美容
 - 法人向け (BtoB)

The screenshot shows the Oricon Customer Satisfaction website. At the top, there is a navigation bar with the Oricon logo and the text "オリコン顧客満足度". Below this, there are several menu items: "調査データ販売", "オリコン顧客満足度* ランキング", and "各種お問い合わせ". The main content area features the heading "学術向けデータ提供" (Academic Data Provision). A large orange banner in the foreground contains the text: "本日協賛企業として参加 → 企業セッションにて プレゼン・ポスターあり" (Participating as a sponsor company today → Presentation and poster at the company session). Below the banner, there are some partially visible text boxes, including "貢献します。" (Contribute).

<https://cs.oricon.co.jp/data/academic/>

ダイエット口コミデータセット ※10月提供開始！

ダイエットカフェ

- 日本最大級のダイエット商品口コミサイト
- 口コミ数16万、商品数8000以上
- ダイエット関連企業とは一切利害関係無し
- ダイエット商材の闇を暴いてます



<https://www.diet-cafe.jp>

→ 企業セッションにて
ポスター展示あり

ダイエット口コミデータセット ※10月提供開始!

ダイエット商品データ



【商品数】 8134

【項目】

- 商品ID
- 商品名
- カテゴリ名
- 商品登録日時
- 口コミ数
- 評価数
- 評価平均点
- 評価内訳
- 特徴

ダイエット口コミデータセット ※10月提供開始！

口コミデータ

★★

2018年3月31日 01:55

広告などを見て、安くなってたので購入しました。
飲んでもあまり変化はありませんでした。
もともとお腹が弱いせいか飲み始めは少し下痢になってしまいました。
約2ヶ月飲みましたが特に効果はありませんでした。
体質によって効果は違うかもですが、個人的にオススメは出来ません。

タグ 副作用_(下痢・便秘等)

終 女性 | 30歳 | 160cm | 使用期間 2ヶ月 | 体重変化 63kg⇒63kg

【口コミ数】 160118

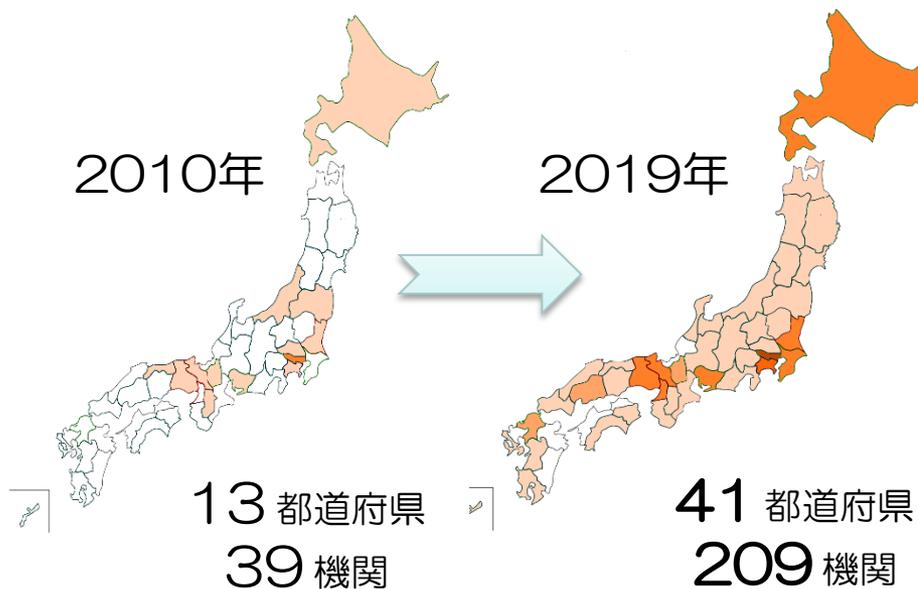
【項目】

- 口コミID
- 投稿者名
- 投稿日時
- 商品ID
- 口コミ本文
- 評価
- 年齢
- 身長
- 使用前体重
- 使用后体重
- 使用期間
- 性別
- タグ(特徴)

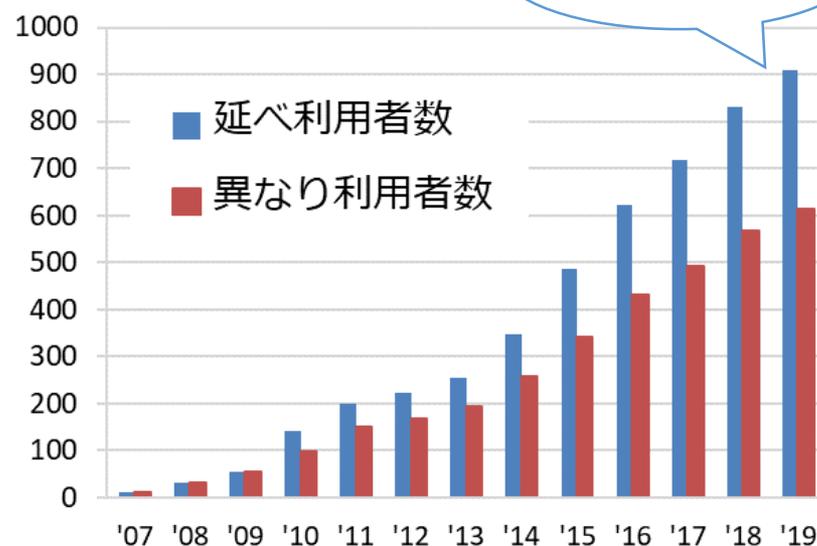
企業データ提供実績：研究室単位（～2019.9）

データ提供先

（ニコニコ，Sansanデータセット除く）

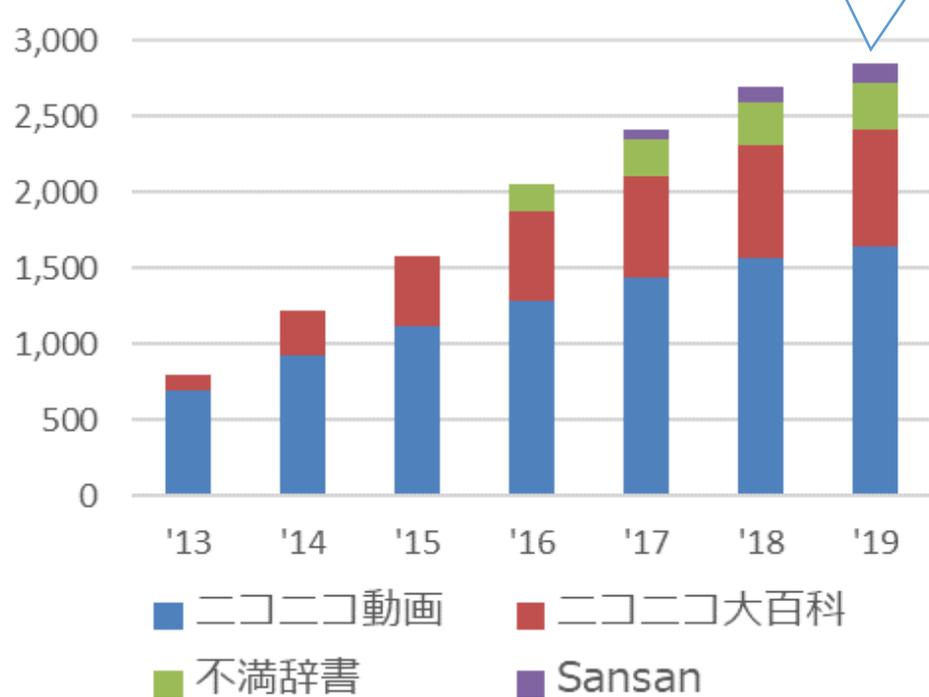


利用者数の推移

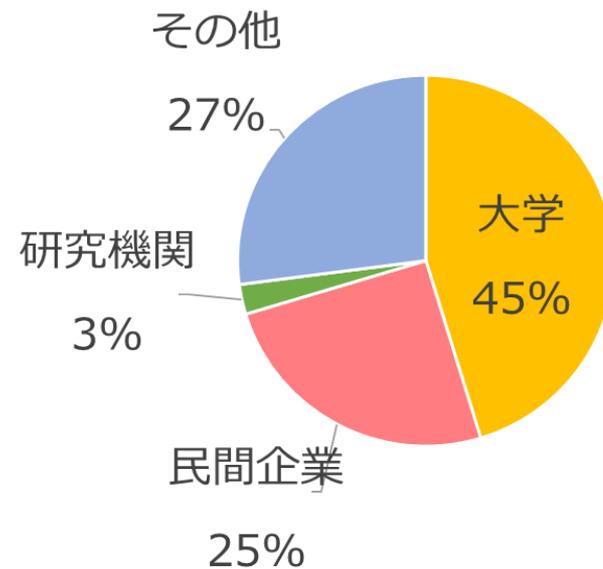


企業データ提供実績：個人単位（～2019.9）

利用申請者数の推移



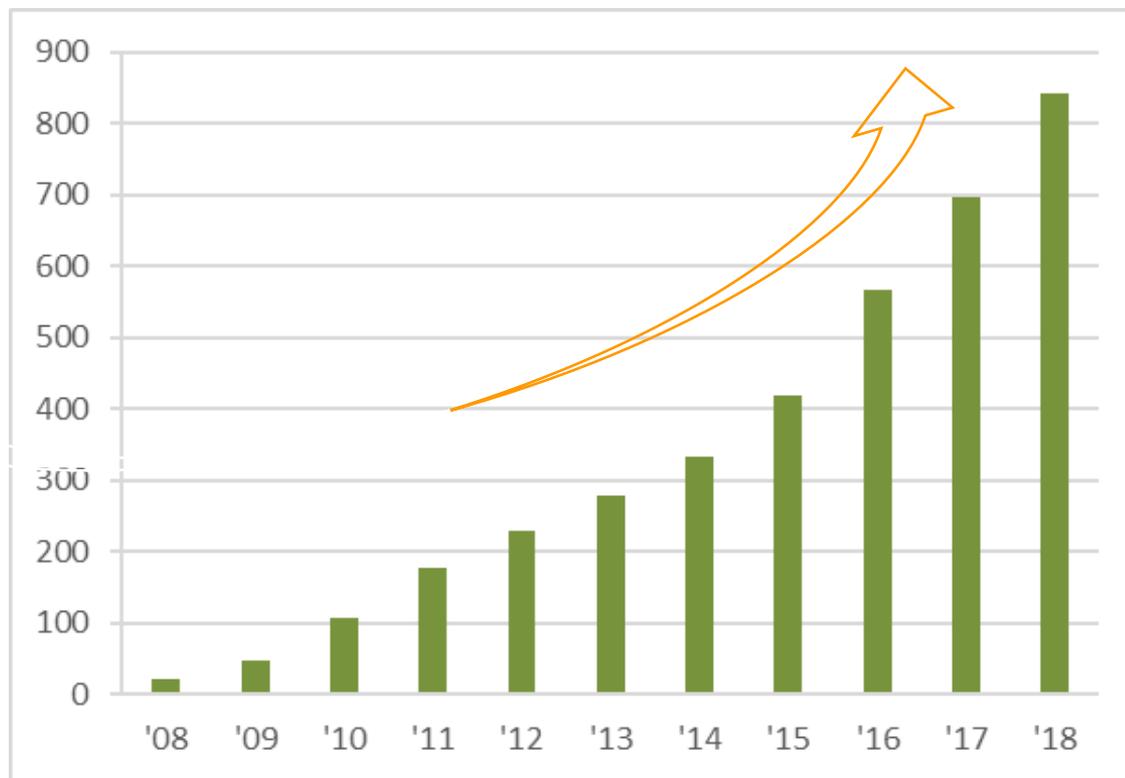
利用者の所属



企業データを利用した研究発表（～2018年度）

研究成果発表数の推移

※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

インデックスリスト

- Yahoo!データセット [226件]
- 楽天データセット [219件]
- ニコニコデータセット [33件]
- リクルートデータセット [9件]
- クックパッドデータセット [137件]
- LIFULL HOME'Sデータセット [25件]
- 不審調査データセット [26件]

アイテムリスト

- ランダムフォレストを用いたカスタマーレビューの半自動分類
- 小暮 枝里子, 宮藤 史哲, 石津 昌平
インテリジェントシステムシンポジウム2017予稿集, 154-157 (2017)
- 定量的分析を用いた自作自演の口コミ判定モデルの構築と評価
- 谷 芽蒔, 藤原 智也, 波多野 昌治
IDRユーザフォーラム2017, (2017)
- レビュー投稿者の満足度を考慮した評判情報の可視化と店舗検索への応用
- 日高 加奈, 豊田 哲也, 大原 則二
IDRユーザフォーラム2017, (2017)
- 価値観モデルを適用した協調フィルタリングによるハイブリッド型推薦手法



<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 callout bubble points to this item. In the center, a list of datasets is shown, with 'オリコンデータセット [1件]' (Oricon Dataset [1 item]) highlighted. A red arrow points from this list item to the right-hand page. The right-hand page displays the details for the '顧客満足度調査データ' (Customer Satisfaction Survey Data). A red box highlights the 'JaLC DOI' field, which contains the value 'info:doi/10.32130/idr.10.1'. A large red callout bubble 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 bottom of the page shows the URL 'https://dsc.repo.nii.ac.jp/'.

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>

<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の記載（※）
- 原稿の事前提出（データセットにより）

【毎年度末に】

- 報告書，論文のリストとコピーの提出
にご協力お願いします！

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/>

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34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43種類
- 映像コーパス（予定含む）
 - 手話データ
 - 多人数会話データ
 - システムとの対話のデータ

IDRの新たな取組

研究者等提供データセットの受入を開始

- 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

The screenshot shows the NII IDR website interface. At the top, there is a navigation bar with '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', and '問い合わせ'. Below this, a sidebar lists various data sets, with '研究者提供データ' (Researcher-provided data) highlighted in a red box. The main content area is titled '大学等研究者提供データセット受入' (University and Researcher Data Set Acceptance), also highlighted in a red box. It includes a date '2019/09/26 作成', an introductory paragraph, '受入条件' (Acceptance Conditions), and '受入手続き' (Acceptance Procedures). The '受入条件' section states that the service is for academic research and that users must agree to terms. The '受入手続き' section lists steps for submission and approval.

English

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

IDR 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へ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、不明な点がございましたら、お気軽にお問い合わせください。

受入手続き

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、送付資料とともに、電子メールの送付ファイルで下記「問い合わせ先」の



映像コーパスの例（1）

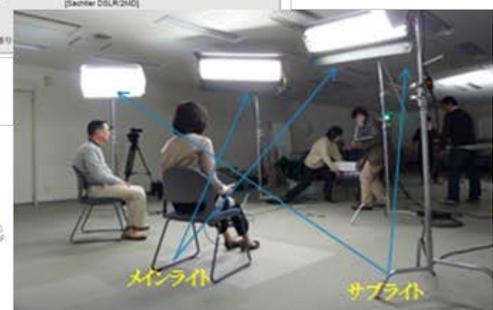
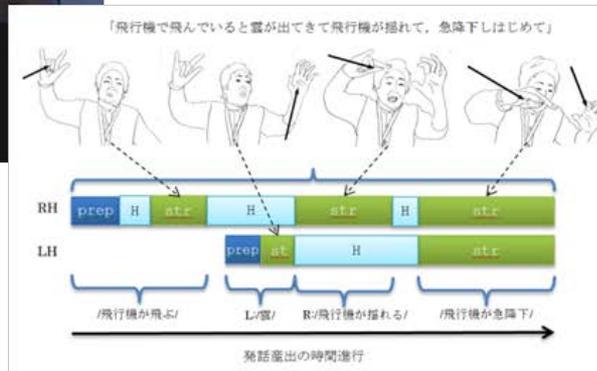
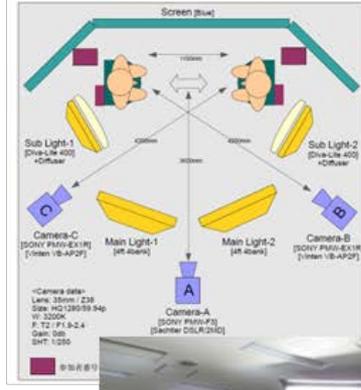
- グループコミュニケーションコーパス ※8月提供開始！



東京電機大学のニュースリリース（2019/2/6）より引用

映像コーパスの例 (2)

- 日本手話話し言葉コーパス (今年度末提供開始予定)



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
- スタートアップセッション
- 招待講演
- 口頭発表セッション

参加企業協賛
情報交換会も
あります！