



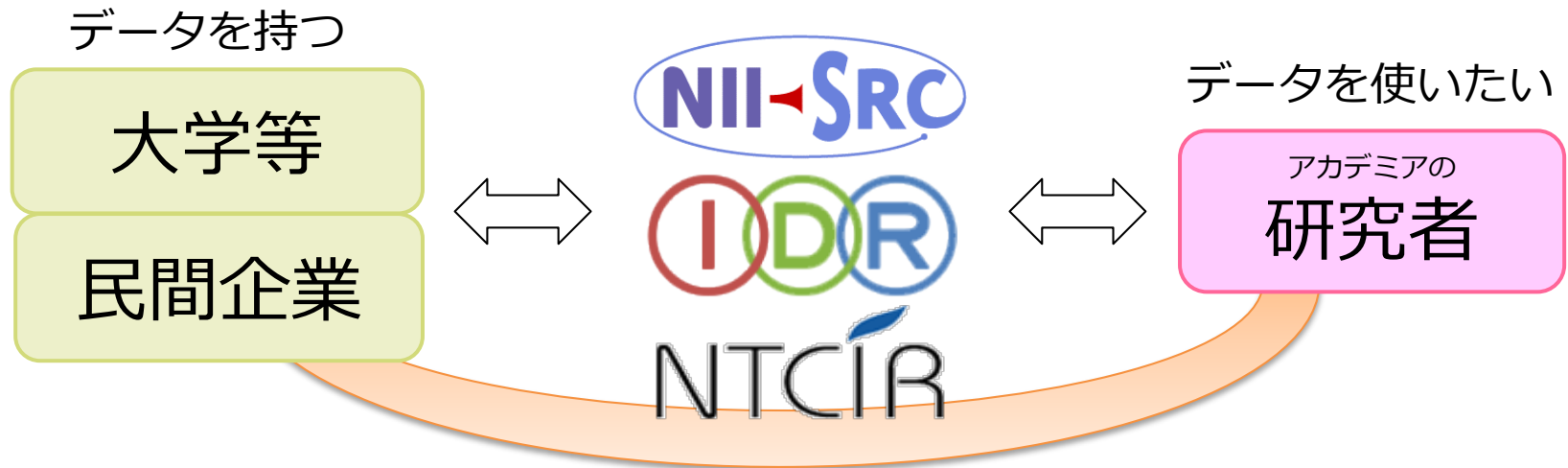
NII-IDRユーザフォーラム2023 ：IDRの活動報告

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
- メルカリデータセット
- 地球の歩き方旅行記データセット

→ 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：みんなの評判口コミデータセット

※2022年8月
提供開始！

マイスタースタジオ社運営の多様な商品・サービスの口コミサイト
「みんなの評判ランキング（みんな評）」への投稿データ



とくめいさん 投稿日：2022.06.10

★★★★☆ 2.00

料金	効果	勧誘・接客	予約
2.00	2.00	2.00	1.00
★★★★☆	★★★★☆	★★★★☆	★★☆☆☆

予約が取れない

予約が全く取れず、なかなか日程も組めないまま半分しかチケットを使っていない状態で契約期間が終了してしまいました。説明では回数保証があるので、購入分のチケット数は必ず使用して頂けます、とのことだったので安心していましたが、実際は返金も残り回数のチケットも全て対応できないとのこと…。説明にありましたか？と疑問に思いましたが諦めました。

参考になった 0

図：脱毛サロンに対する口コミの例

2017年～2022年の投稿

- 約4,700商品・サービス
 - 商品名
 - カテゴリ名
 - サブ評価項目
- 約16万口コミ
 - 投稿者ニックネーム
 - 投稿日時
 - 商品名
 - 口コミタイトル・本文
 - 満足度・サブ評価
 - 「参考になった」の数
 - 気持ちアイコン など

例：メルカリデータセット

※2022年8月提供開始！

メルカリ社が運営するフリマアプリ「メルカリ」への登録データ



図：アプリ画面サンプル

2020年1年間の出品データ

- **商品データ** (100GB)
 - 商品名
 - カテゴリ
 - 商品の説明, 状態
 - 販売価格
 - サイズ, ブランド
 - 発送方法, 発送元地域
 - いいね! の数
 - コメントの数 など
- **コメントデータ** (40GB)
- **画像データ** (3TB)
 - 対象全商品のサムネイル
 - 一週間分のオリジナル

例：地球の歩き方旅行記データセット

※2022年11月
提供開始！

地球の歩き方社運営の「地球の歩き方web」の「旅スケ」に投稿された自由記述の旅行記14,000件とその旅スケジュール9,500件のテキスト



NAIST大内先生ご協力の形態素情報、地名・施設名情報もあり

1 日目 2021 年10 月15 日(金)

...

08:19 - 08:29 郡山駅/東北新幹線

08:29 - 09:40 移動 (電車)

09:40 - 09:46 会津若松駅/磐梯

09:46 - 10:20 若松駅前バス

10:20 - 11:07 移動 (バス)

...

2 日目 2021 年10 月16 日(土)

00:00 - 09:25 庄助の宿 瀧の

09:25 - 09:32 東山温泉入口

09:32 - 09:36 移動 (バス)

...

会津若松へ向かう磐越西線の接続を考慮して選んだやまびこ203号はE5系での運転でした。何度も乗っているE2系よりも座席が広く感じ、快適な移動でした。好天で磐越西線の車窓から磐梯山が良く見えてましたが、山頂だけが雲に覆われていました。

会津若松駅では手作り感のある赤べこのオブジェが出迎えてくれました。これから向かう伊佐須美神社の広告もありました。

伊佐須美神社へ向かう路線バス (永井野行) の出発点として若松駅前バスターミナルを利用しました。

...

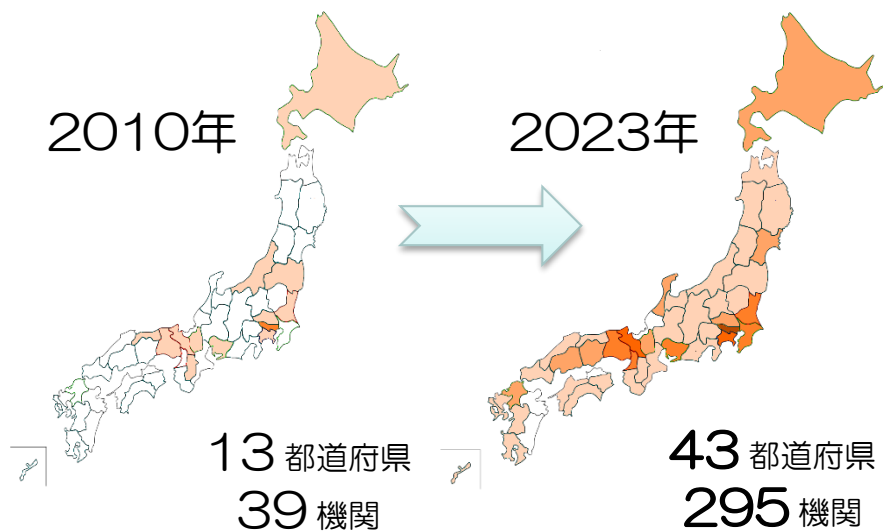
会津若松駅から快速あいづ4号に乗車し、郡山へ向かいました。会津若松は晴れ間がありましたが、山を上るにつれて雲が増えて行き、途中から雨が降り出しました。天気予報通りでしたが、今回の旅行は暖かい2日間で移動時間を除いて雨に降られることがなかったのはラッキーでした。



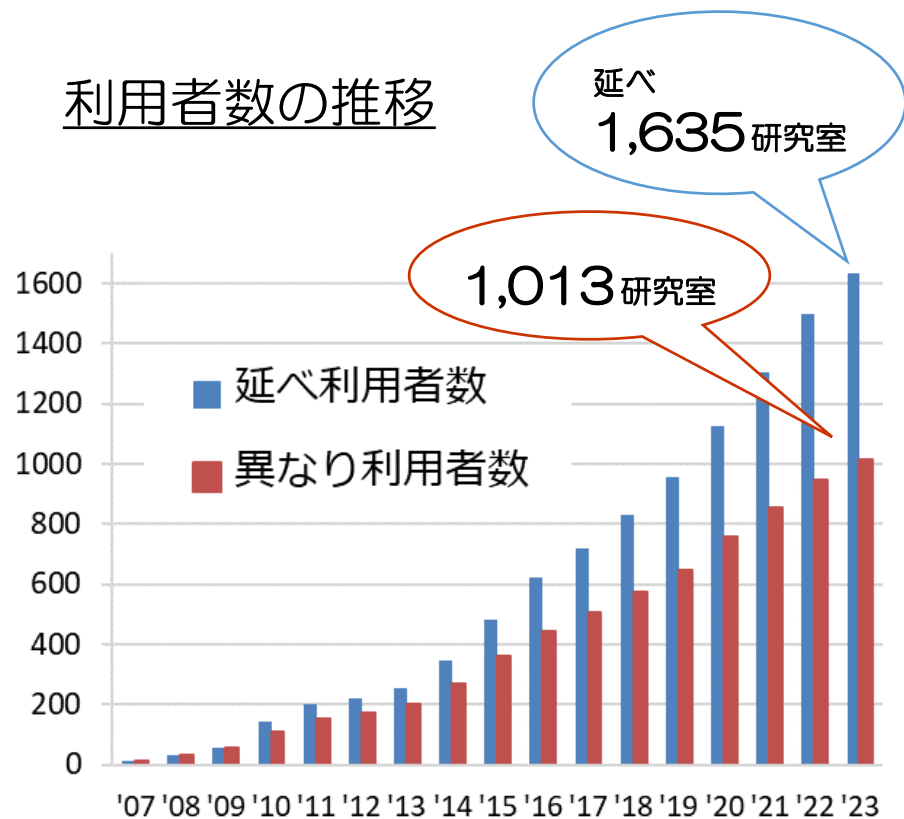
企業データ提供実績：研究室単位（～2023.9）

データ提供先

（ニコニコ， Sansanデータセット除く）



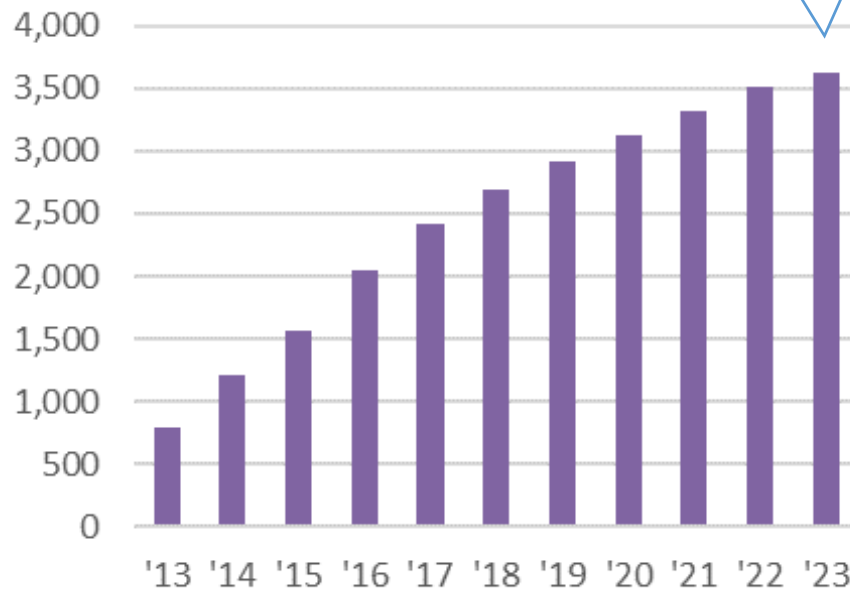
利用者数の推移



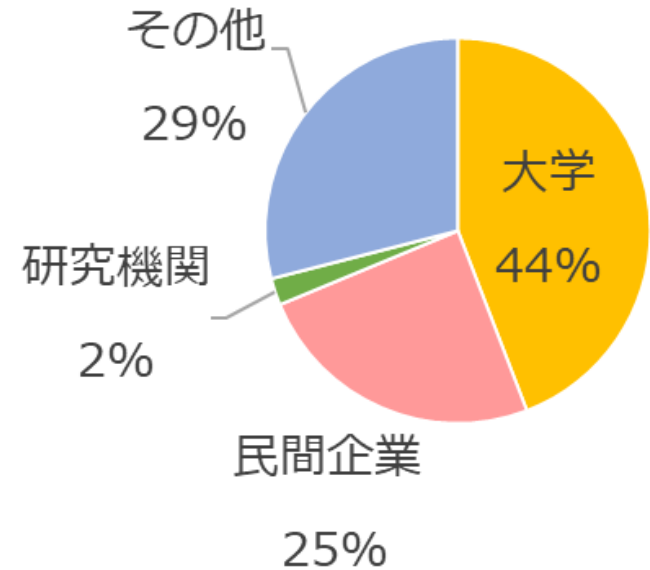
企業データ提供実績：個人単位（～2023.9）

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



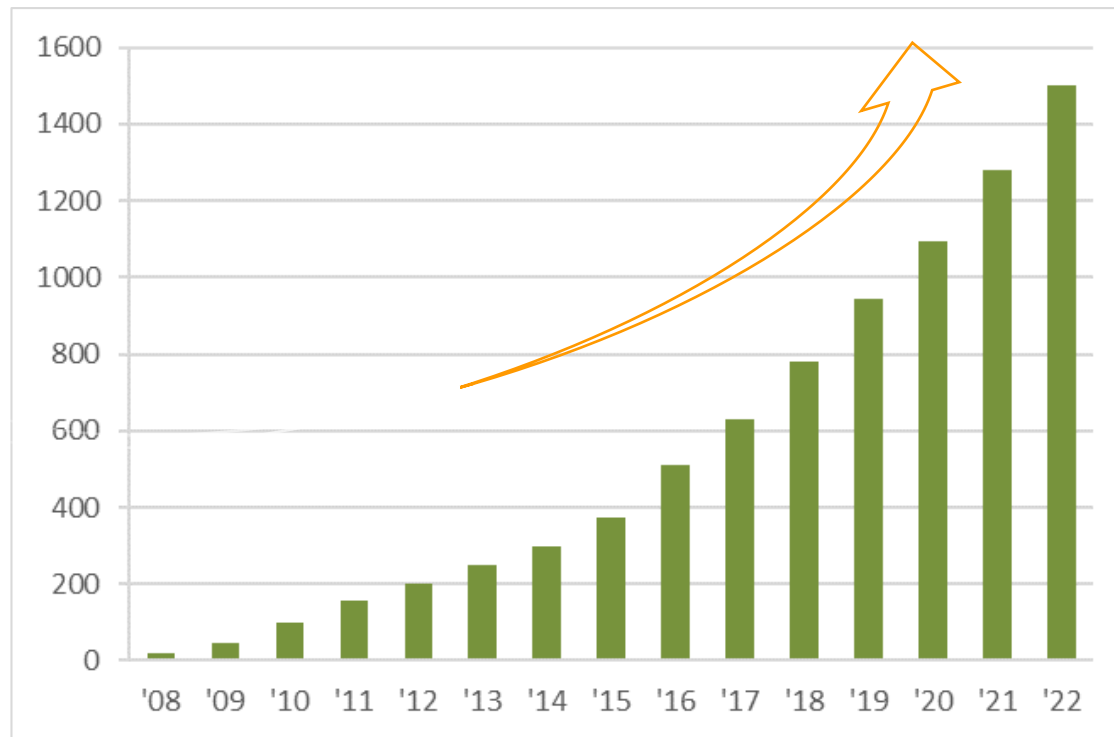
利用者の所属



企業データを利用した研究発表（～2022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 selected. A red box highlights this selection. A red arrow points from this box to the 'オリコンデータセット' entry in the main list. Another red box highlights the 'JaLC DOI' field in the dataset details, 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 top right of the page shows the URL 'Permalink : http://doi.org/10.32130/idr.10.1'.

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

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	株式会社オリコン
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



<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	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



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、 Based on the terms of use of Cookpad
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（課題と正解） 36種類， 文書データ30種類
- 音声コーパス
 - 音声資源コンソーシアムが受入 ～現在52種類
- 研究者等提供データセット
 - IDRが受入 ～現在5種類

2019年～
IDRの新たな取組

研究者等提供データセットの例

• 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)

- 対話エージェントと人との対話の映像と音声データ (1名あたり15分程度, 20代~70代の計約100名)

展示エリアにこれらの
ポスターもあります！

この他に…

- グループコミュニケーションコーパス
- 立命館ARC所蔵浮世絵データベース
- 理研記述問題採点データセット
- 工学院大学 多用途型日本手話言語DB



TC 品					
TC 品 1 品名					
TC 品					

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

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 ユーザーフォーラム 組織 関連リンク お問い合わせ

> HOME > IDRサービス

民間企業提供データ

- 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

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。

サービスの受入の対
ません。
, IDRの能力を超え
い。また、不明な点

NTCIRテストコレクション

音声コーパス

研究者提供データ

- グループコミュニケー
ションコーパス (TDU-
NEDO)

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します。

※ 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「問い合わせ窓口



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
 - データ提供者セッション
 - スタートアップセッション
 - 口頭発表セッション
- (終了後には情報交換会も！)

例年通り
日本DB学会
データチャレンジ
と一体開催！