



データ公開における オープンとクローズの間

水野貴之

国立情報学研究所

情報社会相関研究系

データセット共同利用研究開発センター（DSC）

共有データの重要性



研究の信頼性向上

再現性・透明性の確保
第三者による検証が可能に



研究の活性化・効率化

同じデータで比較・追試が容易
二次分析が新たな発見を生む



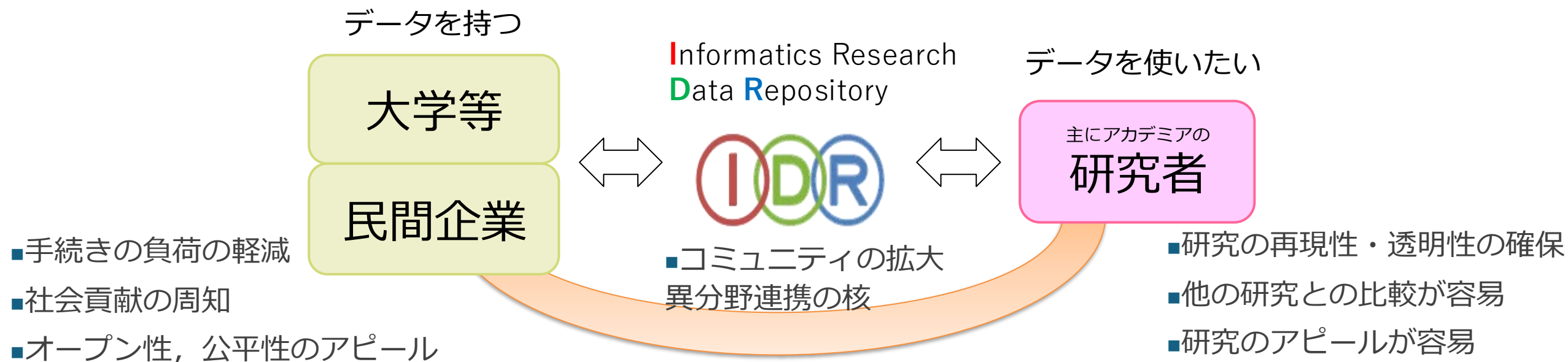
社会的価値の創出

リアルなデータが新たな知見に
政策やサービスへの還元

しかし… 価値の高いデータほど、**個人情報・著作権・企業秘密**などで完全にはオープンにできない

⇒ 適正な管理の下で共有する仕組み = 「オープン」と「クローズ」の間 が必要

データセット共同利用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で取り扱い中のデータセット

- 民間企業提供データセット : 主に秘密情報が理由
→ 16企業より受け入れた34種類
- 研究者等提供データセット : 主に個人情報that理由
→ 1企業・8大学・4機関より受け入れた13種類
- NTCIRテストコレクション
→ 過去のWSで構築されたタスクデータ62種類
- 音声コーパス
→ 音声資源コンソーシアムが受け入れた54種類

提供中の民間企業のデータセット

- 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
- 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
- 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
- REX地価コンテンツデータセット
- ピティナデータセット

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 ユーザーフォーラム 組織 関連リンク お問い合わせ

> HOME

民間企業提供データ

- 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
- 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
- REX地価コンテンツデータセット

NTCIRテストコレクション

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 \(法人利用専用\)](#)」(データセットによっては「[同 \(個人利用専用\)](#)」) 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、民間企業等の方は、「[お問い合わせ](#)」に記載のIDR事務局にご相談ください。

お知らせ

- 「[REX地価コンテンツデータセット](#)」の提供を開始しました。→ [NIIトピックス 2026-02-26 NEW!!](#)
- 「[楽天データセット](#)」を更新しました。→ [NIIトピックス 2025-12-16 NEW!!](#)
- 映像データ収録時の留意事項などを述べた論文が日本言語学会誌に掲載されました。→ 「[映像を含む音声言語資源の共同利用—ディストリビュータの観点から—](#)」 [2025-12-01 NEW!!](#)
- 11月26日(水)に「[IDRユーザーフォーラム2025](#)」を開催しました。 [2025-11-27 UPDATE](#)

履歴

2025-08-18 「[筑波大学 オンラインカウンセリング対話コーパス \(TU-OCDC\)](#)」の提供を開始しました。→ [NIIトピックス](#)

2025-07-17 「[東京科学大学 呼吸波形信号付きマルチモーダル対話コーパス \(BinD\)](#)」の提供を開始しました。→ [NIIトピックス](#)

2025-05-27 「[LIFULL HOME'S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](#)

2025-02-21 「[山梨盲学校 盲ろう児教育実践資料群](#)」の提供を開始しました。→ [NIIトピックス](#)

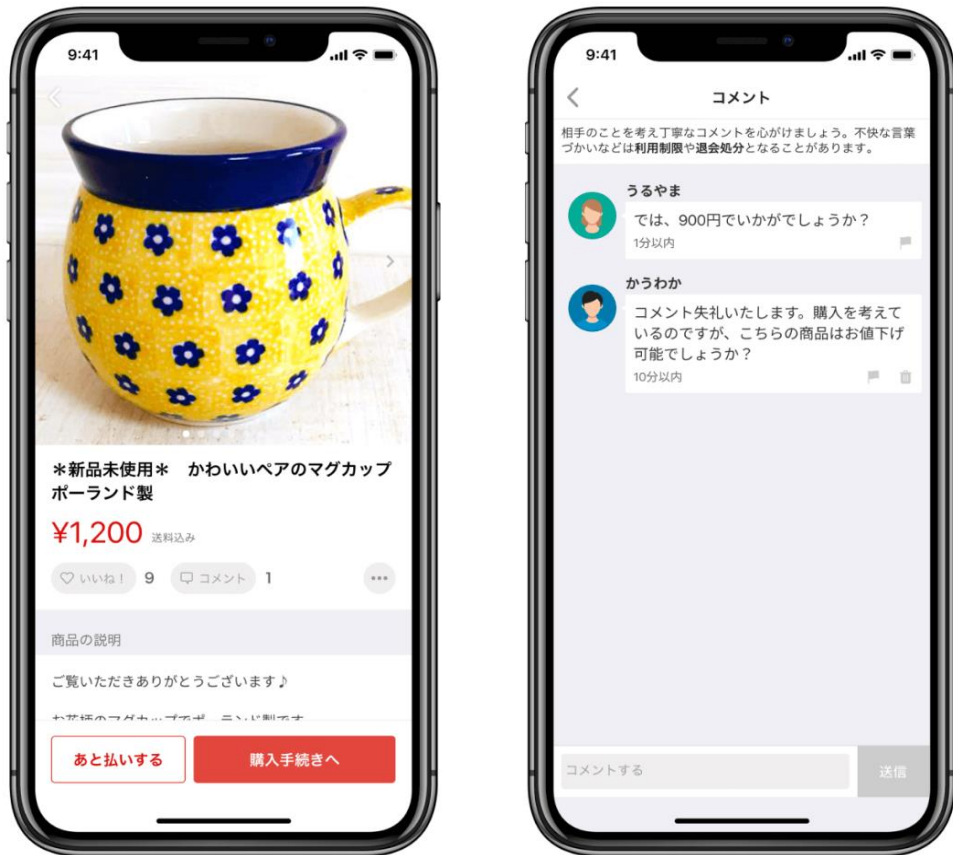
2024-11-25 「[ニコニコデータセット](#)」の「ニコニコ大百科データ」を更新しました。→ [NIIトピックス](#)

2024-11-05 「[立命館ARC所蔵浮世絵データベース](#)」を更新しました。

5

例：メルカリデータセット

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



2020年1年間の出品データ

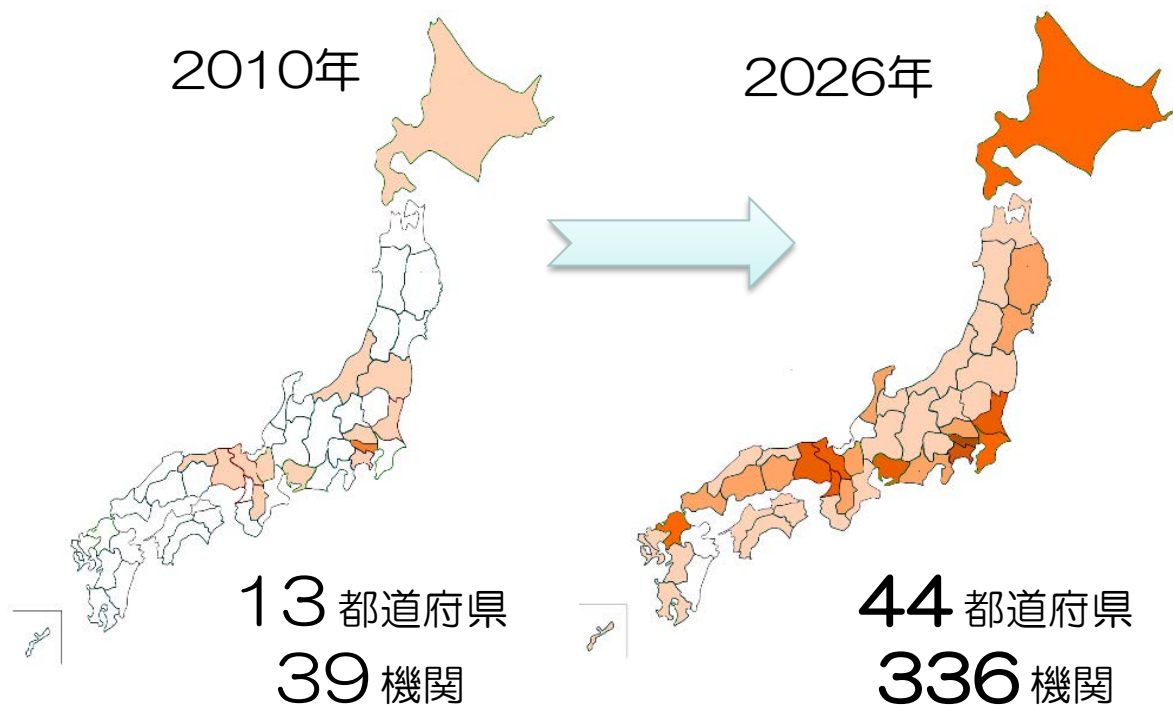
- ・ **商品データ** (100GB)
 - 商品名
 - カテゴリ
 - 商品の説明, 状態
 - 販売価格
 - サイズ, ブランド
 - 発送方法, 発送元地域
 - いいね! の数
 - コメントの数 など
- ・ **コメントデータ** (40GB)
- ・ **画像データ** (3TB)
 - 対象全商品のサムネイル
 - 一週間分のオリジナル

データ提供開始時のプレスリリースより引用
<https://www.nii.ac.jp/news/release/2022/0831.html>

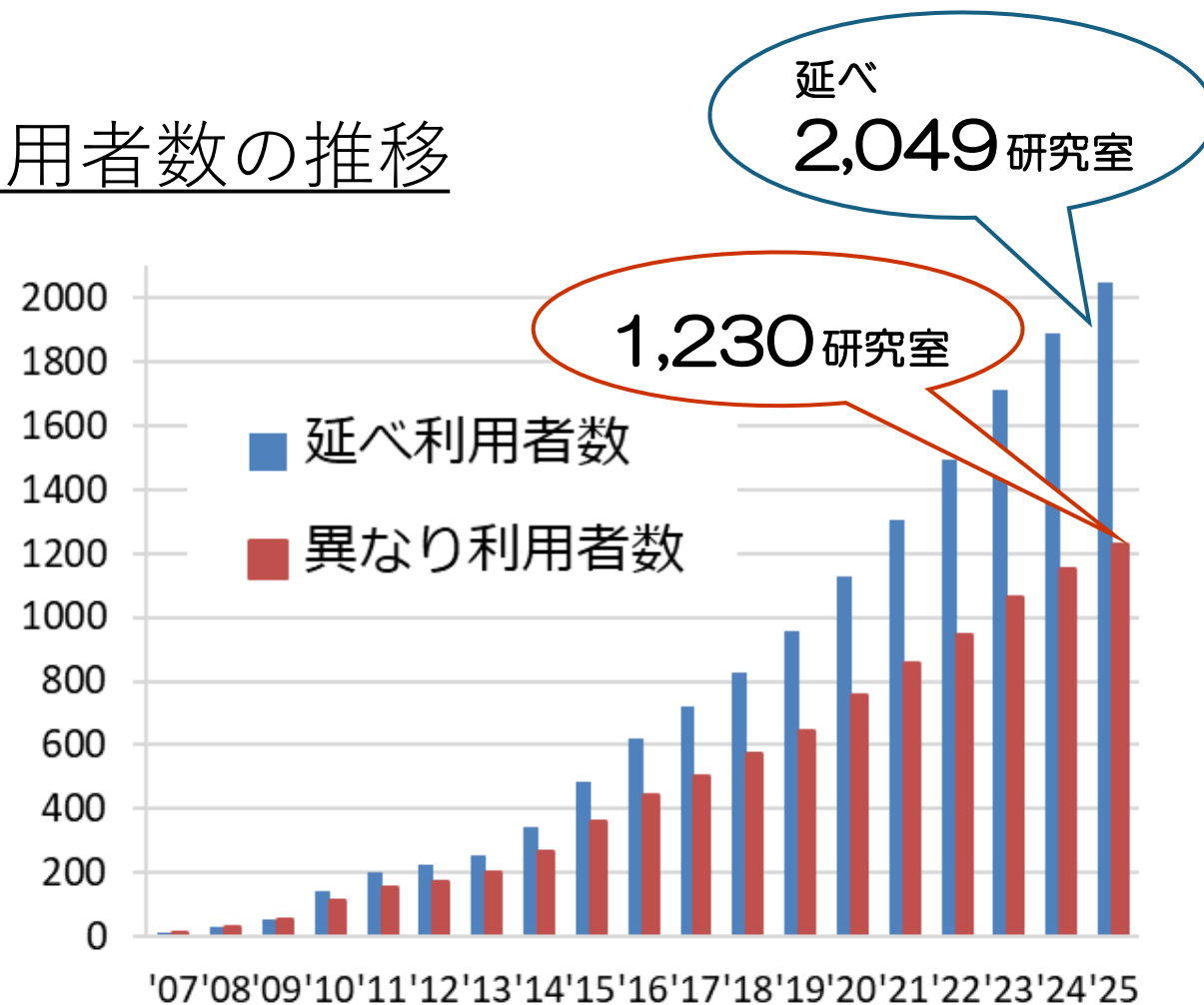
企業データ提供実績（～2026.3）

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



データセットの配布方法

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，…）

- ・ 民間企業の実サービスに基づくデータ
- ・ 会話映像等のセンシティブなデータ
- ・ 研究目的利用に制限されているデータ の提供

① 個人情報等を含むセンシティブなデータ

手続き：メールで利用申請書提出→ 審査 → 同意書（要・公印）の提出

配布：個別URLを生成，パスワードは郵送

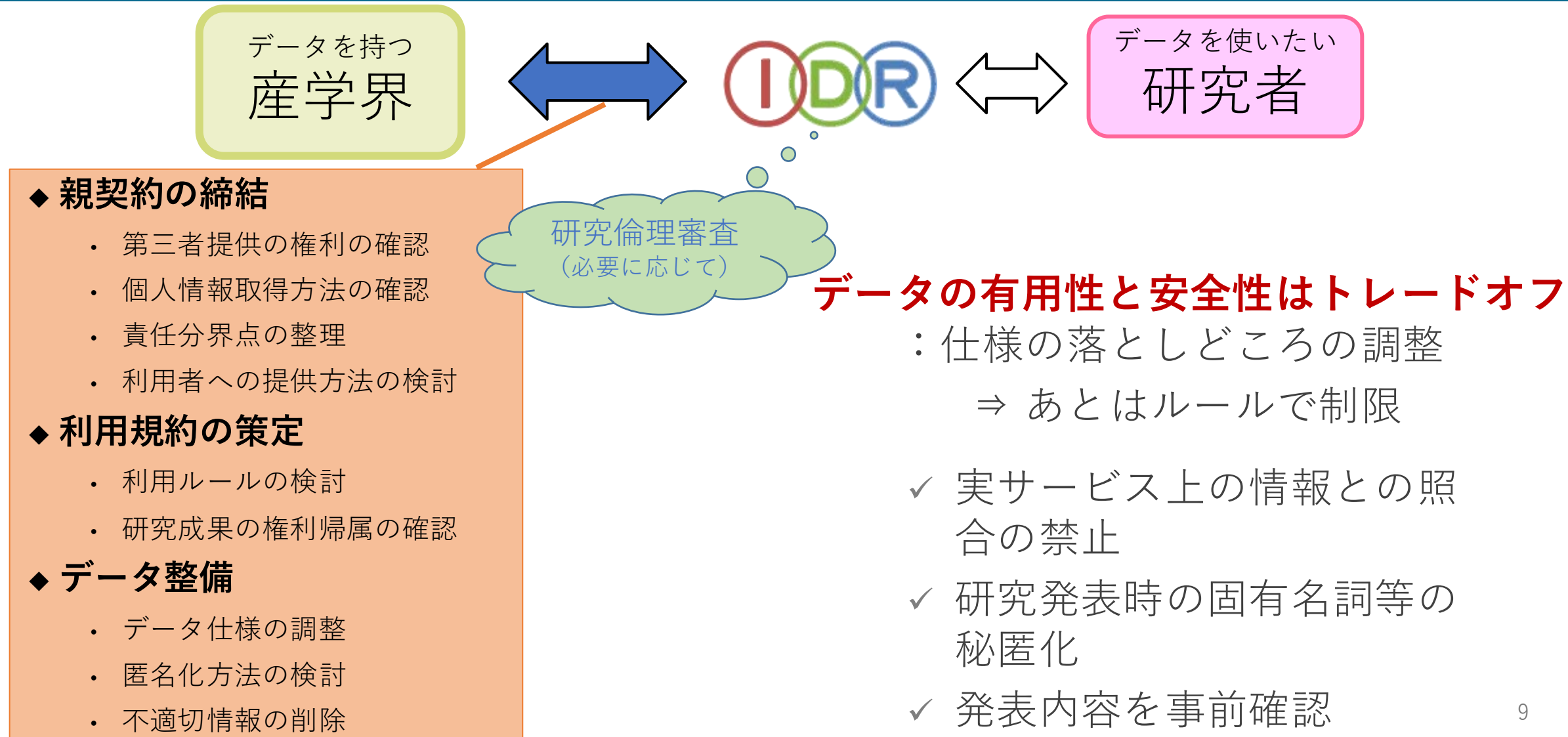
**企業データも
大半はこちら**

② 個人情報等を含まない安全なデータ

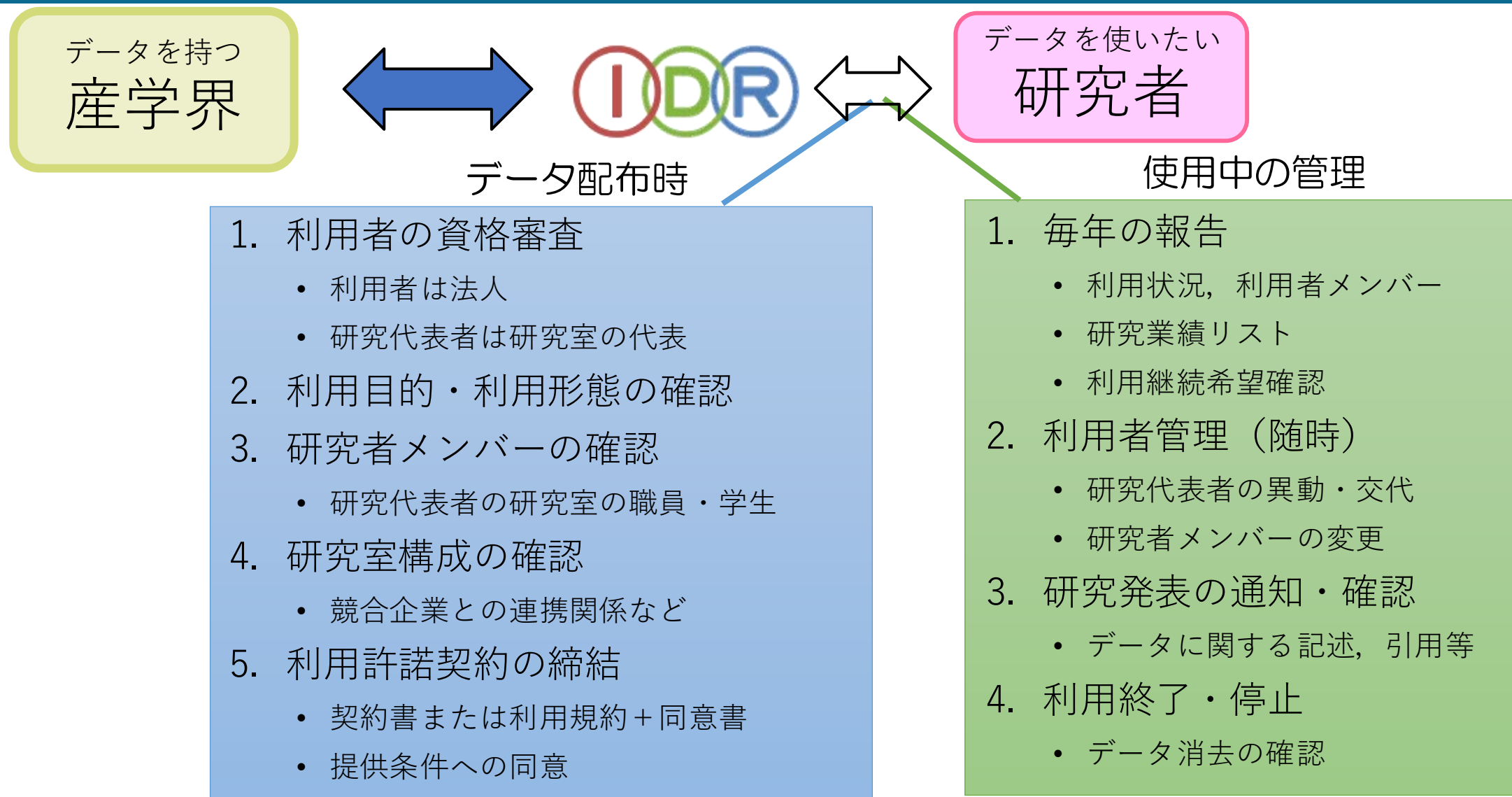
手続き：オンライン申請

配布：オンライン配布

IDRの活動：企業からのデータの受入



IDRの活動：利用者への対応



データDOIも取得

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>

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
提供者	株式会社oricon
配布者	国立情報学研究所
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	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

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

取得収録データの公開を考えている方へのお願い

- 公開/提供の可能性を考慮して，同意書の内容は慎重に！
- サービスの利用規約で第三者提供の同意を確認
- 研究倫理，個人情報，プライバシーへの配慮
 - 場面や話題の選定
 - 収録したデータの処理（ノイズでマスク，仮名化）
- データの権利帰属先のルールは所属機関に依る
 - 事前に提供の可否や条件を要確認
 - 研究倫理審査時にも公開/提供予定の旨を明記

個人情報保護法の除外規定を適用するには，提供先が学術研究機関で，学術研究目的での利用に限られる

└─ ※IDRを通して提供するには原則必要

データの利用者へのお願い

- データの提供を受ける場合は利用ルールを順守
 - 必要な学内の手続きは自らの責任で行う
 - 研究室内でデータも利用者もしっかり管理する
 - ※ 共同研究であってもデータの共有は第三者提供にあたる
 - 公表時には個人情報やプライバシーに配慮する
 - 論文になった時はクレジットを忘れずに

「研究利用だったら何でもあり」ではない

データ作成者への感謝を忘れない！

NIIにてハンドブックも作成

オープンサイエンスのためのデータ管理基盤ハンドブック

オープンサイエンスのための
データ管理基盤ハンドブック

～学術研究者のための“個人情報”の
取扱い方について～

(第1版)

2022年7月27日

国立情報学研究所

オープンサイエンスのためのデータ管理基盤
ハンドブックにかかる検討会 編

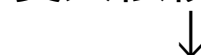
1. ハンドブックの趣旨と想定読者
2. ハンドブックの読み方
3. 定義：対象となる情報とは
4. 定義：対象となる取扱者
5. 定義：その他
6. 研究利用目的での例外の考え方
7. 研究分野毎の留意点
8. 研究における個人情報の利用形式の検討
9. 研究の各側面における個人情報取扱の注意点：研究計画の策定
10. 研究の各側面における個人情報取扱の注意点：研究における個人情報の取得
11. 研究の各側面における個人情報取扱の注意点：研究における個人データの管理
12. 研究の各側面における個人情報取扱の注意点：個人情報の共有
13. 研究の各側面における個人情報取扱の注意点：保有個人データ・保有個人情報の取扱い
14. 講ずべき安全管理措置

IDRを通した共有をご検討の場合

- 研究者等提供データ
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

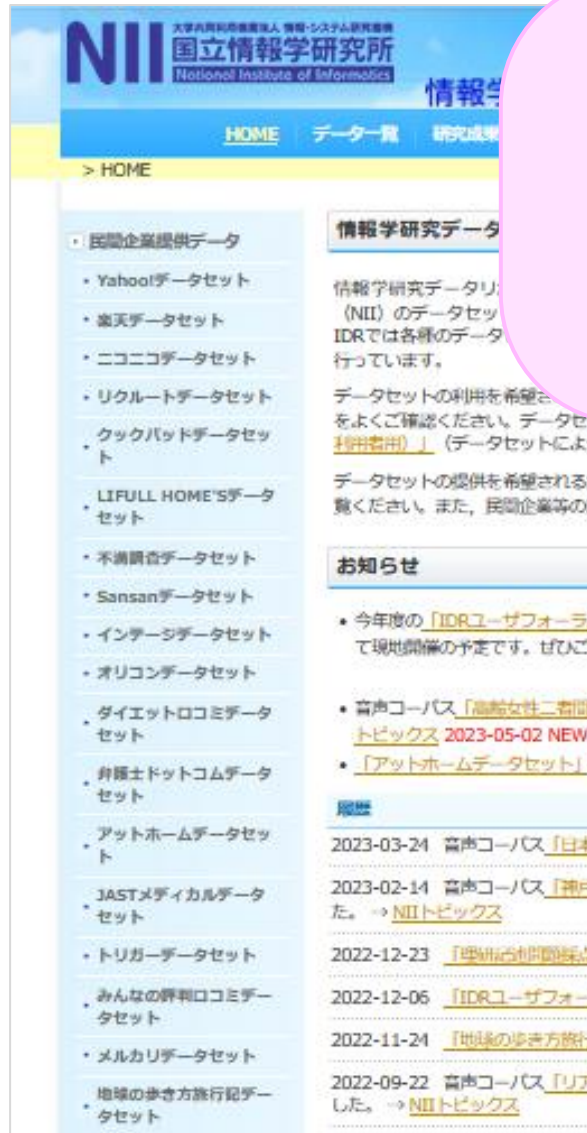
※民間企業データは
個別対応

The screenshot shows the NII (National Institute of Informatics) website. The header includes the NII logo and the text '国立情報学研究所 National Institute of Informatics'. The main navigation bar has links for '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 the navigation bar, there is a section titled '情報学研究データリポジトリ' (Information Science Research Data Repository) with a sub-header 'IDR'. A red box highlights the '大学等研究者提供データセット受入' (University and Researcher Data Set Acceptance) link. The page content includes a date '2019/09/26 作成' (Created 2019/09/26), a paragraph explaining the IDR service, a section titled '受入条件' (Acceptance Conditions) with detailed text, and a section titled '受入手続き' (Acceptance Procedure) with a numbered list of steps. The left sidebar contains a list of data sets, including '民間企業提供データ' (Private Company Provided Data), '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', '弁護士ドットコムデータセット', '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トラッキングデータ (TDU-NEDO)', and '立命館ARC所蔵浮世絵データベース'.

詳しい受入要件・手続きについては…

- ・ 個人情報等を含むデータ
- ・ 研究目的利用に制限したいデータ

⇒ **idr@nii.ac.jp** 宛てに
お気軽にご相談下さい！



ご清聴ありがとうございました

IDRに興味を持って下さった方は …

メール : idr@nii.ac.jp

NII IDR

検索

