



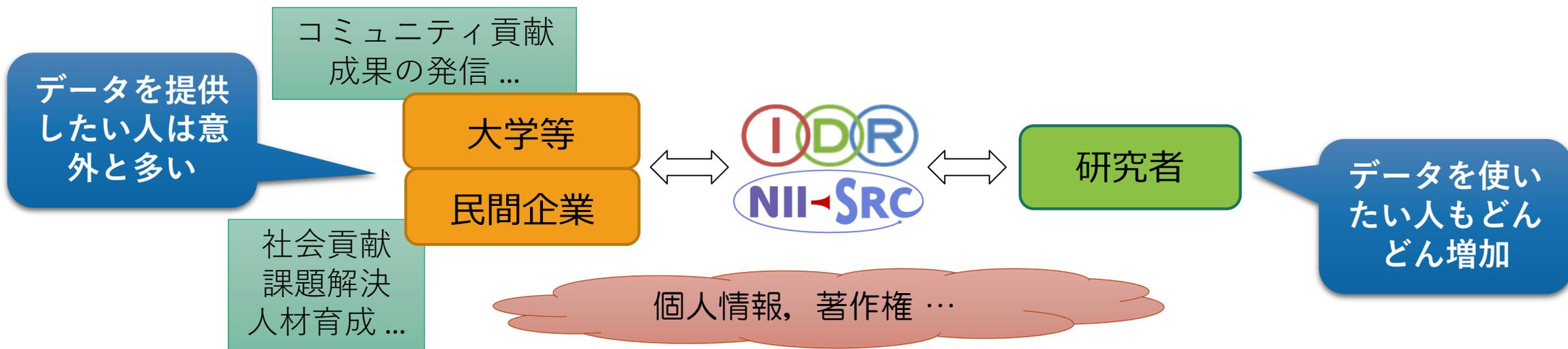
研究者の自発的データ提供の 促進に向けた取組 — NII-IDRの活動を通して —

大山敬三

国立情報学研究所 (NII)

データセット共同利用研究開発センター (DSC)

産学界や大学等のデータを学術研究目的で研究者に提供



「オープン」にできないリアルなデータを
契約に基づき適正な管理の下で利用可能に

- ★ 安全・適正なデータ提供のノウハウを提供
- ★ 企業や研究者にとって面倒な契約等の手続を代行

⇒ 提供のハードルを下げる



提供時・使用中の管理（IDRが代行）

提供時

1. 利用者の資格審査
 - 利用者は法人
 - 研究代表者は研究室の代表
2. 利用目的・利用形態の確認
3. 研究者メンバーの確認
 - 研究代表者の研究室の職員・学生
4. 研究室構成の確認
 - 競合企業との連携関係など
5. 利用許諾契約の締結
 - 契約書または規約＋同意書
 - 提供条件への同意

使用中

1. 毎年の報告
 - 利用状況，利用者メンバー
 - 研究業績リスト
 - 利用継続希望確認
2. 利用者管理（随時）
 - 研究代表者の異動・交代
 - 研究者メンバーの変更（オプション）
 - 研究室構成の変更（オプション）
3. 研究発表の通知・確認
 - データに関する記述，引用等（オプション）
4. 利用終了・停止
 - 利用報告のない者の追跡，停止通知

基本的にはいずれもデータセット提供者のポリシーに基づく

スケールさせるにはシステムの支援が必要

データ提供のインセンティブ：研究成果論文リストの公開

データ提供の学術貢献を可視化

研究発表

DSC Reference Portal

データセット共有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学術情報や論文リストを公開しています。

インデックスリスト

- 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]
- MEAL HOME データセット [4件]
- 不潔調査データセット [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
- 公費 成子, 喜島 実由, 石津 真生
- インフルエンザウイルス感染シミュレーション (2017) 学報, 154-157 (2017)
- 定額分析を用いた自作自演のロコモティブシンドロームの検出と評価
- 益彦, 蓮生 智也, 渡辺 舞
- IDRユーザフォーラム2017, (2017)
- レビュー投稿者の社会性を考慮した評判情報の可視化と読者検索への応用
- 日高 加菜, 豊田 直也, 大塚 剛三
- IDRユーザフォーラム2017, (2017)
- 機械学習モデルを用いた協調フィルタリングによるハイブリッド推薦手法

DSC Reference Portal
<https://dsc.repo.nii.ac.jp/>



CiNii Research
<https://cir.nii.ac.jp/ja>

CiNii Research

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 2015-02-24
- 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
Web Site 研究データあり 参考文献1件
- Mining online activity data to understand food consumption behavior: Asian fish sauce among Japanese consumer
Mitsutoshi, Nakano, Hiroaki, Sato, Toshihiro, Watanabe, Katsumi, Takano, Yoshimasa, Sagane Food Nutrition 2018
DOI Web Site 研究データあり 参考文献1件
- 既存レシピの調理手順を用いたアレンジレシピ作成支援システムの提案
長屋, 健, 佐藤, 哲司 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 2018



IDRで提供中のデータセット

企業提供データセット

1. Yahoo! データセット
2. 楽天データセット
3. ニコニコデータセット
4. リクルートデータセット
5. クックパッドデータセット
6. LIFULL HOME'Sデータセット
7. 不満調査データセット
8. Sansanデータセット
9. インテージデータセット
10. オリコンデータセット
11. ダイエットロコミデータセット
12. 弁護士ドットコムデータセット

大学等研究者提供データセット

1. グループコミュニケーションコーパス（東京電機大学）
2. 立命館ARC所蔵浮世絵データベース（立命館大学）
3. 三重大 日本語地図課題対話コーパス（三重大学）（SRC）
4. 理研記述問題採点データセット（理化学研究所）
5. 大阪大学 マルチモーダル対話コーパス（大阪大学）

提供の申出があったにもかかわらず、提供に至らなかったデータの方が多い

提供の障害となる要因（データ提供者）

企業提供データセット

- データ取得条件
 - 利用・提供権限（契約条件）
 - 著作権，著作者人格権
 - 意外に穴がある
 - 個人情報
 - 適正な取得？
 - 使用目的（学術研究？）
 - 第三者提供（提供条件？）
- 不適切データへの対策
 - プライバシー侵害，誹謗・中傷
 - 検出・削除の取組・体制

※ 法的リスク，炎上リスク

大学等研究者提供データセット

- 関係する要素
 - データのオーナーシップ
 - 研究倫理審査，個人情報
- 提供を困難にする状況
 - 知財関連規則の変遷
 - 帰属の研究資金による場合分け，留保
 - 研究者の異動
 - 複数の研究者・機関の関与
 - 機関間の規則・制度の差異
 - 複数の研究資金

※ トレースし，不備はリカバー
（二重の困難）

⇒ データ構築までのトレースがシステム化されるとうれしい