

オープンデータ ～日本版データリポジトリの可能性～

武田英明

takeda@nii.ac.jp

国立情報学研究所

デジタル化以前の研究者

研究と執筆

論文



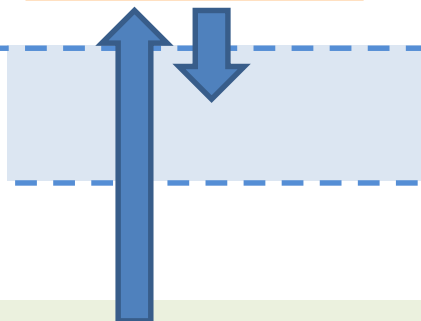
文献調査



論文執筆

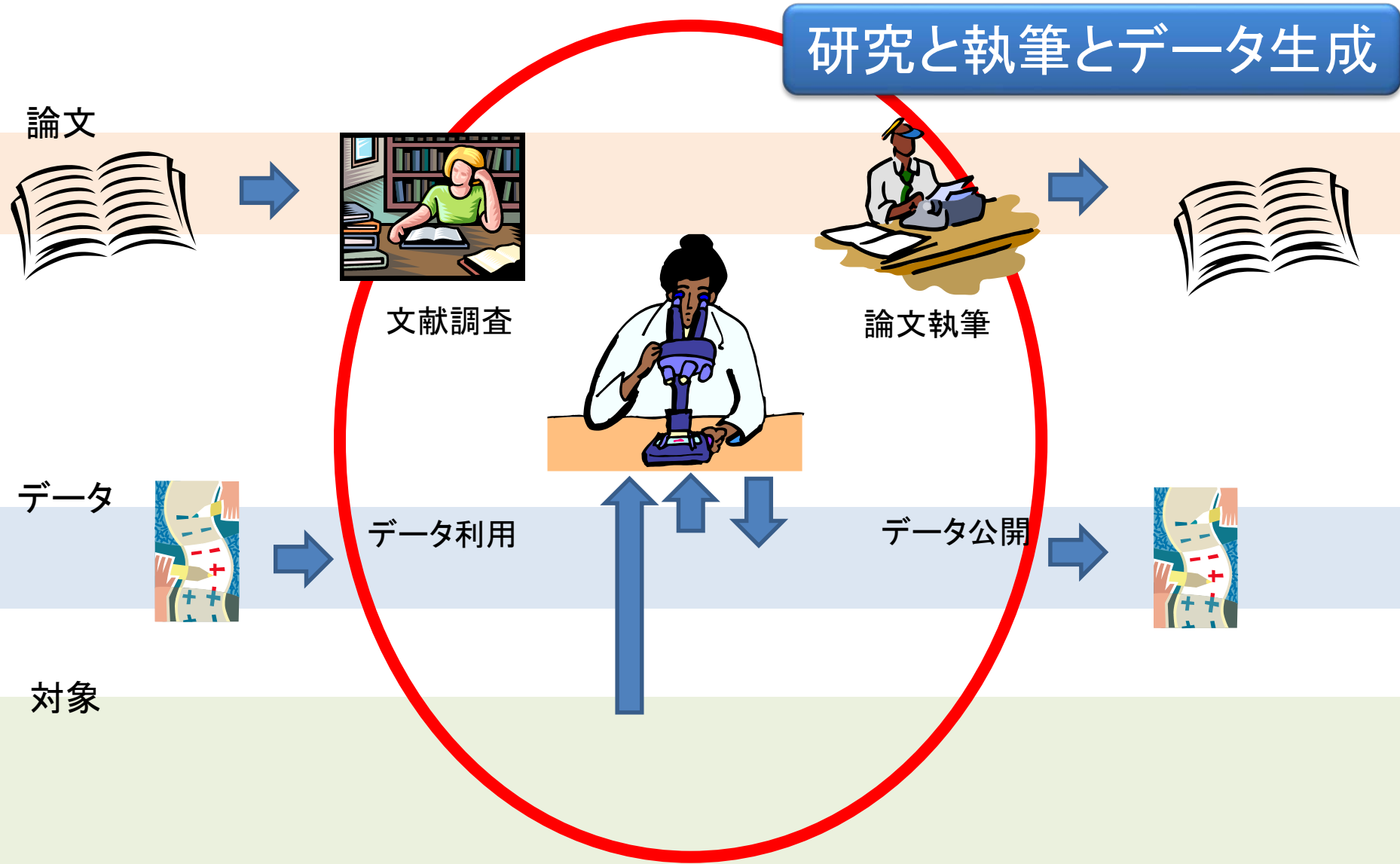


データ



対象

現在の研究者



今後の研究者

論文・データの
一体化

研究とデータ生成

論文



データ利用

データ公開

データ



対象

研究＝データのサプライチェーン

研究成果のデータ化

{ 研究成果としてのデータ公開
論文のデータ化 }

- 研究成果データとの一体化
 - データ付録
 - データ論文
- 論文自体がデータとして扱われる
 - 大量論文のデジタル処理
 - 人間のための論文からコンピュータのための論文へ

なぜ研究データをオープンにするか

- 社会的要請
 - 社会での成果の共有、知識の共有
 - 公的資金の公共性のため
 - 例：研究資金助成機関のオープンデータポリシー
- 研究のオープン性
 - 研究の発展性のため
 - “巨人の肩に乗る”
 - 再現性担保のため

科学のタイプと再現性

- 第1の科学：理論科学
 - 演繹
- 第2の科学：実験科学
 - 再実験による仮説検証
- 第3の科学：シミュレーション科学
 - 再シミュレーションによる検証？
- 第4の科学：データ中心科学
 - ？？（再プロセスによる検証？）

科学のタイプと再現性

- 第3や第4の科学(計算科学)においては論文は研究の成果ではなく、研究の紹介でしかない。
- では研究の成果はどうやって公開？
 - プログラム、プロセス(ワークフロー)、データの公開
 - 論文との連携

研究成果のデータ化

{ 研究成果としてのデータ公開
論文のデータ化 }

- 研究成果データとの一体化
 - データ付録
 - データ論文
- 論文自体がデータとして扱われる
 - 大量論文のデジタル処理
 - 人間のための論文からコンピュータのための論文へ

Research Data Alliance (RDA)

- 2013年3月発足
- データ駆動型のイノベーションを加速することが目的
- 社会的、組織的、技術的インフラ構築を目指す
 - データ共有、交換の障害を減らす
 - 協調的なグローバル・データ・インフラをつくる
- ヨーロッパとアメリカが主体で立ち上げ
- これまで3回の会合



ERDP-2012-01

Title: Long-term hydrochemical monitoring in Oyasan Experimental Forest Watershed comprised of two small forested watersheds of Japanese cedar and Japanese cypress

Authors: Urakawa *et. al.*

Accession Number: ERDP-2012-01

Accepted: Jan 2012

Published on: Ecological Research (2012) 27: 245. doi: 10.1007/s11284-012-0926-8

Click to see [Metadata](#).

Data files are stored in [JaLTER Database](#).

News

Update: ERDP-2012-01

May 27, 2014

ERDP-2013-02 available

Dec 20, 2013

Update: ERDP-2012-02, ERDP-2012-03

Jun 12, 2013

ERDP-2013-01 available

May 08, 2013

Data Catalog Search

[Home](#)

Data Set Citation:

When using this data, please cite the data package:

Toda H.

Long-term hydrochemical monitoring in Oyasan Experimental Forest Watershed comprised of two small forested watersheds of Japanese cedar and Japanese cypress

ERDP-2012-01.3.5 (<http://db.cger.nies.go.jp/JaLTER/metacat/metacat/ERDP-2012-01.3.5/default>)

General Information:

Title: Long-term hydrochemical monitoring in Oyasan Experimental Forest Watershed comprised of two small forested watersheds of Japanese cedar and Japanese cypress
(ja) スギ・ヒノキが植栽された2つの人工林小流域からなる大谷山試験流域の水文水質データ
Identifier: ERDP-2012-01.3.5

Data Table, Image, and Other Data Details:

Metadata Ecological Metadata Language (EML) File
download:
Data Table: precipitation_discharge ([View Metadata](#) | [Download File](#))
Data Table: waterchemistry ([View Metadata](#) | [Download File](#))

論文のデータ化とデータ公開

- 論文のデータ化
 - 論文データベースとデータによる評価
 - Citation Index, Altmetrics
 - 識別子とメタデータ
 - DOI
 - 論文からデータへ (DataCite)
 - ORCID
 - 研究者へのID
 - FundRef
 - 研究資金機関へのID
 - Linked Open Data (LOD)
- データ公開とデータ共有基盤
 - データ・リポジトリ

DOI (Digital Object Identifier)

- 識別子(DOI name)からデジタルオブジェクトが存在するURIに変換するサービス
- International DOI Foundation(IDF)が運営するサービス(1998年～)
- 元々は出版社が識別子を共有するにつくった制度であるが、現在はより広範に様々なデジタルオブジェクトの識別子を提供する制度に。
- 分散管理
 - IDFは登録業務を行う団体(Registration Agency, RA)に委譲
 - CrossRefは一つのRA

DOI

- デジタルオブジェクトの識別子とその解決を行うサービス
 - 識別子の付与
 - デジタルオブジェクトの識別子から存在するURIへの変換 (Handle system)
- DOIの存在意味
 - 永続する管理主体
 - 運営ルール
 - システム拡張
- DOIのメリット
- サービス提供
 - 持続性
 - 識別子が有効性を維持
 - 一貫性
 - 識別子が指すものの“保証”

CrossRef

- STM出版における引用関係を管理する仕組み
- 2000年～
- DOIのRA
- 機能：
 - DOIのメタデータ(書誌データ、引用関係)の管理
 - DOI登録
 - メタデータを用いたサービス
 - 書誌検索、引用検索
 - Reverse look up: メタデータを検索して対応するDOIを返す
 - この機能を用いて、出版社は個々の引用文献にDOIを付与

CrossRef

- デジタルオブジェクトの識別子とその解決を行うサービス(Handle system)
 - 識別子の付与
 - デジタルオブジェクトの識別子から存在するURIへの変換
- DOIの存在意味(Handle system単独と比較して)
 - 永続する管理主体、運営ルール、システム拡張
- DOIのメリット(Handle system単独と比較して)
 - サービス提供、持続性、一貫性
- CrossRef
 - より高度のサービス提供
 - 書誌、引用管理
 - Reverse look up
 - 持続性、一貫性の強化
 - 特定領域のデータクオリティの維持

Japan Link Center (JaLC)

- 国立国会図書館(NDL)、国立情報学研究所(NII)、科学技術振興機構(JST)、物質・材料研究機構(NIMS) による共同運営として設立
- 機能
 - IDFのRAとしてメタデータ付きのDOIの発行管理
 - CrossRefのメンバーとしてCrossRef DOI登録仲介
- ミッション
 - 国内文献のDOI付与の促進
 - 国内ニーズによる多様なDOIの発行管理

DataCite

- データを参照するための仕組み
- データセットにDOIをつける

[Home](#) [Members](#) [FAQs](#) [Services](#) [Resources](#) [Events](#) [Contact us](#)



Helping you to find,
access, and reuse data

DataCite

DataCite Business Models Principles are online

Published by Frauke Ziedorn on 31 October 2012 - 1:08pm

DataCite's Business Models Principles document is now available for [download](#). Intended for DataCite members and their clients, the document looks at DataCite member and client responsibilities, as well as best practices.

[Read more](#)

DataCite Services

Published by Frauke Ziedorn on 18 October 2012 - 1:32pm

DataCite offers various services supporting the DOI registration and discovery of research data. These services include the Metadata Store, a statistics portal, the OAI-PMH Provider and many more. Please have a look at our [Services website](#) for more information!

[Read more](#)

Re-scheduled Open Hours for Datacenters

Published by Frauke Ziedorn on 10 October 2012 - 7:48am

DataCite will host open hours for data centres on Tuesday, 16th October 2012. 5pm Berlin time. Please look [here](#) for your local time.

Why cite data?

What is DataCite?

What do we do?

DOI resolver

Resolve a DOI string (e.g. 10.5284/1000164) by entering it below:

Metadata Search

**PANGAEA®**

Data Publisher for Earth & Environmental Science

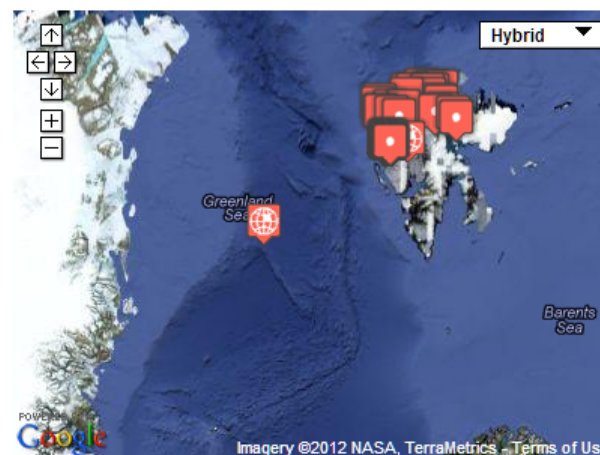
Not logged in (log in or sign up)

Always quote citation when using data!

Data Description
[Show Map](#)
[Google Earth](#)
[RIS](#)
[BibTeX](#)

Citation: Ushio, S et al. (2008): Oceanographic data in the Greenland Sea and around Svalbard in 1991-1993. doi:10.1594/PANGAEA.701498, Supplement to: Ushio, Shuki; Ito, Hajime; Ono, Nobuo (1998): Oceanographic data in the Greenland Sea and around Svalbard in 1991-1993. *NIPR Arctic Data Reports, National Institute of Polar Research, Tokyo*, 3, 46 pp, hdl:10013/epic.30872.d001

Abstract: The Arctic Ocean is connected with the North Atlantic Ocean by the Fram Strait between Greenland and Svalbard. The strait is located in the northern part of the Greenland Sea. In the eastern part of the strait, warm saline water flows northward as the West Spitsbergen Current; while in the western part, cold less-saline water flows southward as the East Greenland Current. The northwestern part of the Greenland Sea is normally covered with sea ice even in summer. Furthermore, this region is regarded as a major area where the Arctic sea ice is discharged into mid latitude oceans. Thus, this area plays an important role in heat and salt exchange processes in the Arctic marine system.



The reveal exchange processes of water masses and ocean-atmosphere interaction in high-latitude oceans, a number of international research programs have been focused on the Greenland Sea and its surrounding waters. As one of the international Arctic research programs, oceanographic studies have been executed in cooperation with the Norsk Polarinstitutt and other institutes under the leadership of the National Institute of Polar Research since 1991. Japanese scientists have been carrying out field observations in and around Svalbard. The observations include not only physical measurements but also biological surveys. This report presents physical oceanographic data obtained in the Greenland Sea in 1992 and 1993, and data around Svalbard from 1991 to 1993.

Project(s): **National Polar Research Programme of Japan (NIPR)** 🔍

Coverage: Median Latitude: 79.212347 * Median Longitude: 13.300383 * South-bound Latitude: 74.150000 * West-bound Longitude: -13.350000 * North-bound Latitude: 80.183333 * East-bound Longitude: 20.066667

Date/Time Start: 1991-09-10T00:00:00 * Date/Time End: 1993-08-29T10:47:00

Event(s): **KF1991_A0** 🔍 * Latitude: 78.991667 * Longitude: 11.966667 * Date/Time: 1991-09-10T00:00:00 * Location: Kongsfjorden, Svalbard 🔍 * Campaign: KF1991 🔍 * Basis: smal boat 🔍 * Device: CTD/Rosette 🔍

KF1991_A1 🔍 * Latitude: 78.998333 * Longitude: 11.988333 * Date/Time: 1991-09-12T07:55:00 * Location: Kongsfjorden, Svalbard 🔍 * Campaign: KF1991 🔍 * Basis: smal boat 🔍 * Device: CTD/Rosette 🔍

KF1991_A2 🔍 * Latitude: 79.003333 * Longitude: 12.053333 * Date/Time: 1991-09-12T07:30:00 * Location: Kongsfjorden, Svalbard 🔍 * Campaign: KF1991 🔍 * Basis: smal boat 🔍 * Device: CTD/Rosette 🔍

DataCite Metadata Properties

- Mandatory Properties

<i>ID</i>	<i>Property</i>
1	Identifier (with type attribute)
2	Creator (with type and name identifier attributes)
3	Title (with optional type attribute)
4	Publisher
5	PublicationYear

- Optional Properties

<i>ID</i>	<i>Property</i>
6	Subject (with scheme attribute)
7	Contributor (with type and name identifier attributes)
8	Date (with type attribute)
9	Language
10	ResourceType (with description attribute)
11	AlternateIdentifier (with type attribute)
12	RelatedIdentifier (with type and relation type attributes)
13	Size
14	Format
15	Version

figshare

[Browse](#)[Upload](#)[Sign up](#)[Login](#)

Table 1: Relationship between Mendeley Readership and Altmetric score

Table 1: Relationship between Mendeley Readership and Altmetric score			
Item	ALL	OA	NON-OA
Readership on MIE group	67.7	67.5	68.0
Readership on Mendeley	144.0	159.5	103.3
Altmetric score	14.8	16.9	9.4
readers on Mendeley	45.6	40.0	60.6
readers on Connotea	1.5	2.0	0.4
readers on CiteULike	5.6	6.9	2.1
Tweeted by	7.9	9.2	4.3
Blogged by	1.1	1.4	0.6
Reddited by	0.1	0.1	0.0
On Facebook pages	0.3	0.4	0.1
Mentioned in Google+ posts	0.3	0.2	0.4
Picked up by news outlets	0.1	0.0	0.1

17
views

0
shares

cites
coming
soon

Published on 16 Feb 2014 - 07:32 (GMT)
Filesize is 65.51 KB

Categories

- Computer Engineering

Authors

Keita Bando

Tags

- Mendeley
- altmetrics

License (what's this?)

CC-BY

Export

- Export to RefWorks
- Export to BibTeX

[Enlarge](#)[Download](#)

Share this:



0



0



0



Cite this:

Bando, Keita (2014): Table 1: Relationship between Mendeley
Readership and Altmetric score. figshare.
<http://dx.doi.org/10.6084/m9.figshare.9385689>

研究者の識別子

- 人の識別子 ≠ デジタルオブジェクトの識別子
 - 人 ≠ デジタルオブジェクト
- 実体としての「研究者」の多面性
 - 論文、書籍の著者
 - 研究プロジェクトの参加者
 - 大学・研究機関のメンバー
 - ...
- 研究者の識別子への要求仕様
 - ユニーク性
 - 多面性

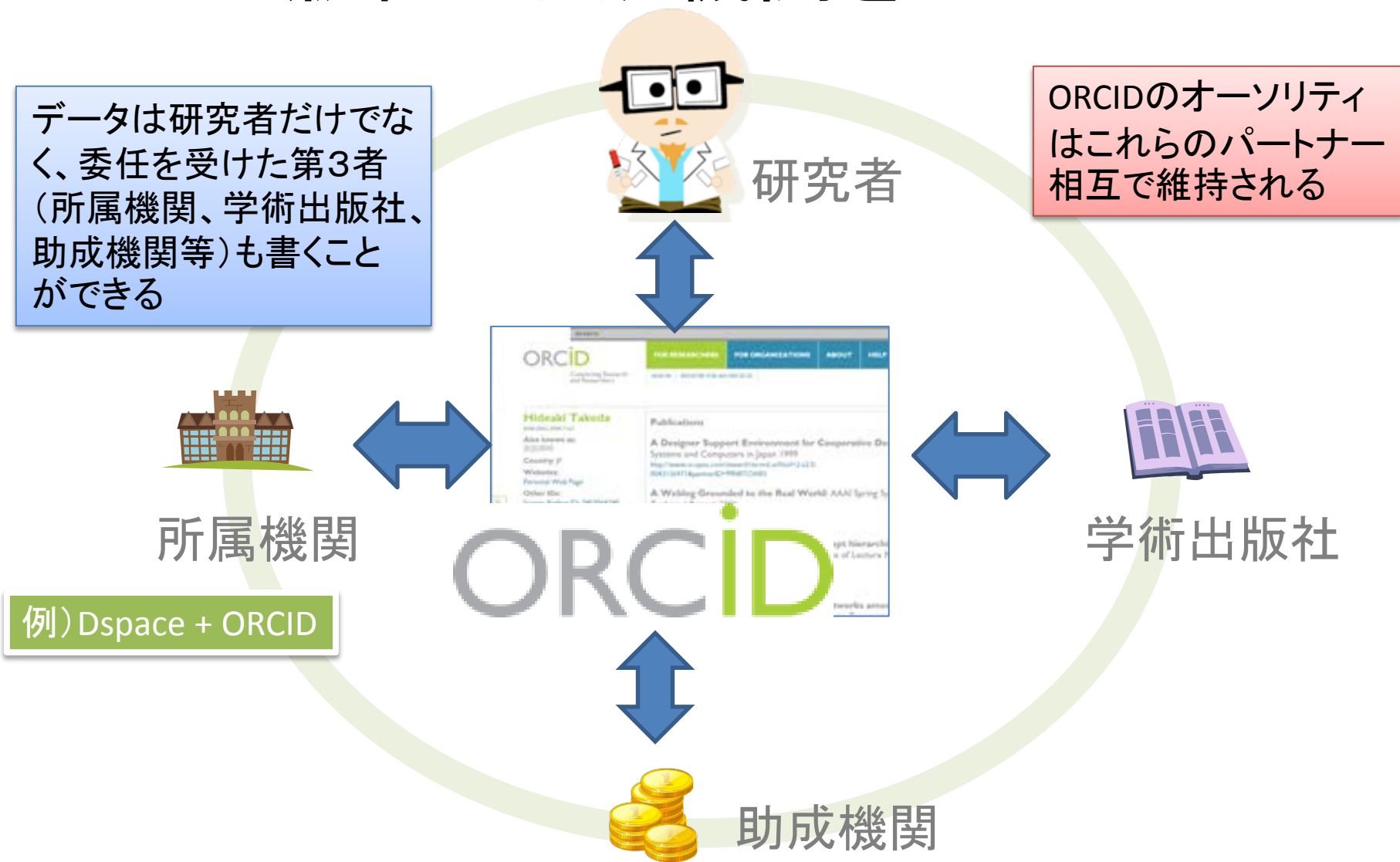
ORCID



(Open Researcher and Contributor Identifier)

- 著者や研究寄与者を一意に同定するための識別子
- NPO 2011年～
- 出版社、大学、助成機関などのメンバーからなる
- サービス開始: 2012年10月～
- 利用法:
 - 論文投稿時
 - 論文の著者欄
 - 大学の業績管理
 - ...

ORCIDは研究者と所属機関と学術出版社と助成機関をつなぐ



研究データ・リポジトリ

- 様々な形態が想定される
 - 基本機能：
 - 識別子管理(コンテンツ(DOI),研究者(ORCID)、Fundなど)
 - メタデータの管理と共有
- 構築方法
 - 1. 機関リポジトリの発展形
 - 2. データ・カタログサイト
 - 3. 分野特化型サイト

次のステップへ(1)

- 機関リポジトリの拡張としてのデータリポジトリ
 - 論文の代わりにデータ、データセット
 - データは機関リポジトリの中あるいは別のところ
- メリット
 - 慣れたシステムを使える
 - 論文と統合的に扱える
- デメリット
 - 巨大データなどはおけない
 - データ用メタデータと論文用メタデータの違い

次のステップへ(2)

- データカタログサイト
 - データセットのメタデータを格納、公開するサイト
 - データは基本は外部(小規模なら内部でも)
 - 例 ckan、Socrata
 - メリット
 - ソフトウェアの実績
 - オープンデータサイトとの連携
 - デメリット
 - 機関リポジトリとは非連携
 - 学術データ特有の各種IDなどには非対応

次のステップへ(3)

- 分野固有リポジトリ
 - 各分野でのデータサービスをリポジトリとして
 - 例 : NASA Data Repository
 - メリット
 - それぞれの要求に合わせたデータ格納、アクセス
 - それぞれの要求に合わせたメタデータデータ
 - デメリット
 - メタデータが非連携
 - 独自ソフトウェア

日本版データ・リポジトリへの課題

- 誰が大学の担当か
 - 図書館、計算機センター、管理部
- 誰が全体の旗振り役か
 - NII、JST、JSPS、文科省、NICT
- どんなデータをターゲットにするのか
 - 大きさは？ : ビッグデータ、中小データ、個別のデータ
 - 形態は？ : データベース、構造化データ、非構造化データ
 - 更新頻度は？ : 累積的、定期的、静的
 - メタデータは？ : 詳細(分野特有)、簡略(分野横断)
 - 利用、共有の度合いは？ : 分野横断、分野内

日本版データ・リポジトリへの課題

- 提案

- Small Start

- 機関リポジトリからのスタート

- 例) Weko+金沢大学での取り組み、学術資源リポジトリ協議会

- JaLC+DataCiteによるDOI付与

- Big Picture

- 関係者会議の創設

まとめ

- 研究成果のデータ化の進行
 - 研究データの公開
 - 論文のデータ化
- 日本版データ・リポジトリへの課題
 - 多様なリポジトリのあり方
 - 多様なデータ・リポジトリの関係者
- 私の提案
 - Small Start: 機関リポジトリからの実践、DOI付与
 - Big Picture: 関係者会議

お知らせ

- ORCID Outreach Meeting @ Tokyo
 - ORCIDの活動を広くお知らせするイベント
 - 2014年11月4日 NIIにて開催
 - （図書館総合展の前日）

