

# 機関リポジトリとDOI ～研究データ流通を支える情報基盤とは～

武田英明

[takeda@nii.ac.jp](mailto:takeda@nii.ac.jp)

ORCID: 0000-0002-2909-7163

- 「論文」の未来 ～人間可読から機械可読へ～
- 研究データ流通を支える情報基盤とは

- 「論文」の未来 ～人間可読から機械可読へ～
- 研究データ流通を支える情報基盤とは

# デジタル化以前の研究者

研究と執筆

論文



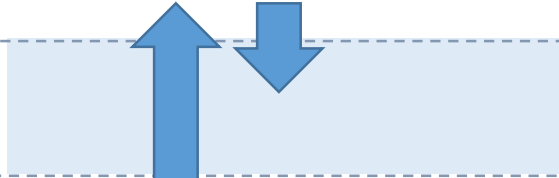
文献調査



論文執筆



データ



対象

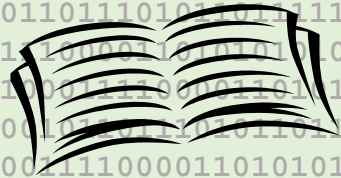


# 今後の研究者

論文・データの一体化

研究成果＝データ生成

論文



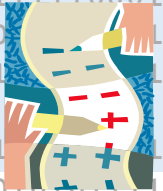
データ利用



データ公開



データ



研究＝データのサプライチェーン

対象

# 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

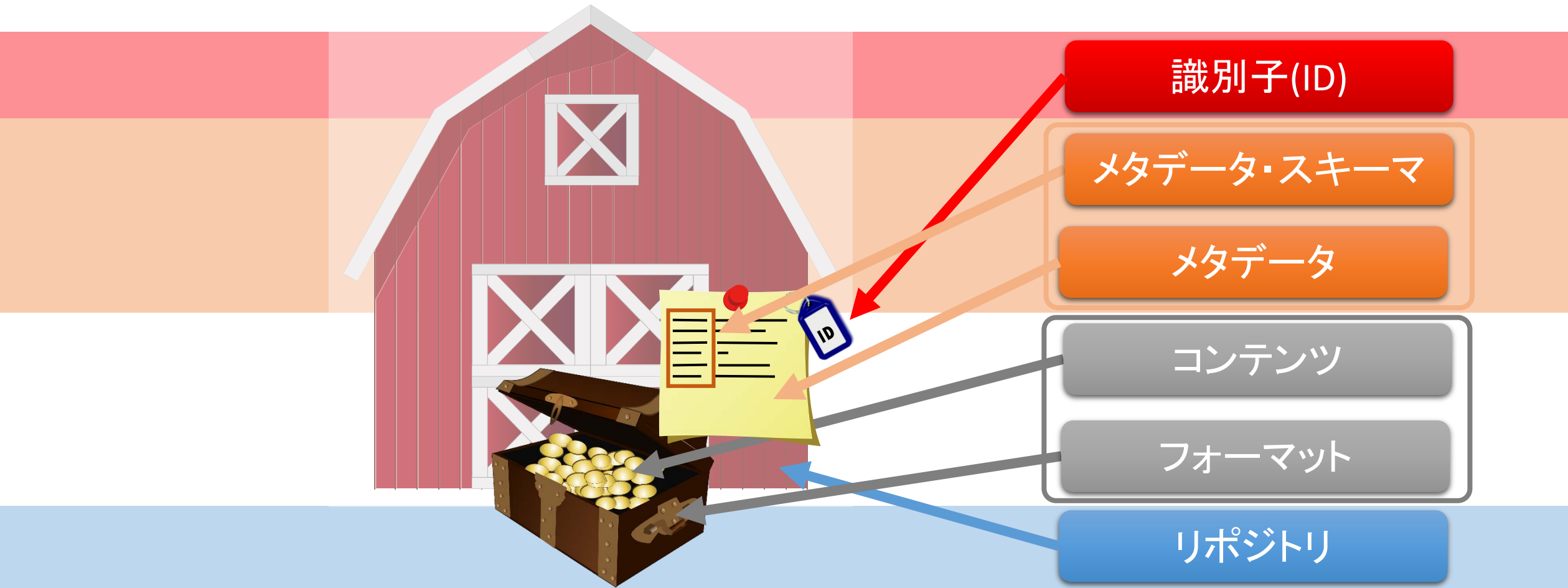
- 第1の科学：理論科学
  - 演繹
- 第2の科学：実験科学
  - 再実験による仮説検証
- 第3の科学：シミュレーション科学
  - 再シミュレーションによる検証？
- 第4の科学：データ中心科学
  - ?? (再プロセスによる検証?)

# 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

- 「論文」の未来 ～人間可読から機械可読へ～
- 研究データ流通を支える情報基盤とは

# 研究データ流通を支える情報基盤のレイヤー



# 研究データ流通を支える情報基盤のレイヤー

識別子(ID)

管理組織・体制、ID連携

DOI

ORCID

FundRe  
f

メタデータ・スキーマ

スキーマ記述言語、スキーマの収集・共有、スキーマ変換

組織

DataCite

CrossRef

JaLC

スキーマ

Dublin Core

DCAT

システム

CKAN

技術

Linked Data

メタデータ

メタデータ記述言語、メタデータ収集・検索、メタデータ変換

コンテンツ

フォーマット

リポジトリ

システム開発、コミュニティ

# 研究データ流通を支える情報基盤のレイヤー

識別子(ID)

管理組織・体制、ID連携

メタデータ・スキーマ

スキーマ記述言語、スキーマの収集・共有、スキーマ変換

メタデータ

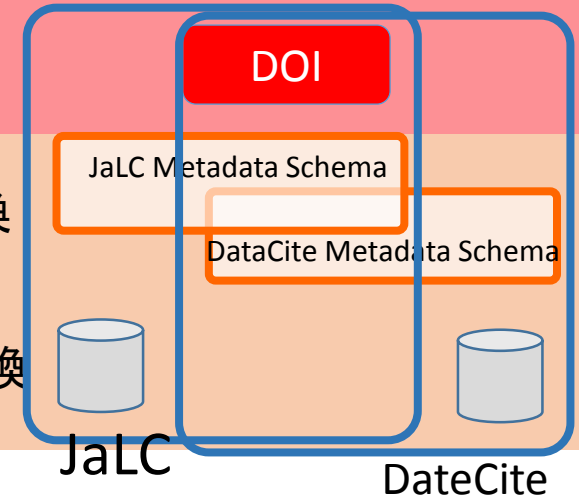
メタデータ記述言語、メタデータ収集・検索、メタデータ変換

コンテンツ

フォーマット

リポジトリ

システム開発、コミュニティ



# DOIの役割

- 解決可能(resolvable)、持続的(persistent)、相互運用可能な(interoperable)リンクを提供
  - 解決可能: 標準化したID構文 + handle systemによるマッピング
  - 持続的:
    - 技術的インフラ: レジストリDBの運用、
    - 社会的インフラ: 登録機関(RA)への義務
  - 相互運用可能: データモデル

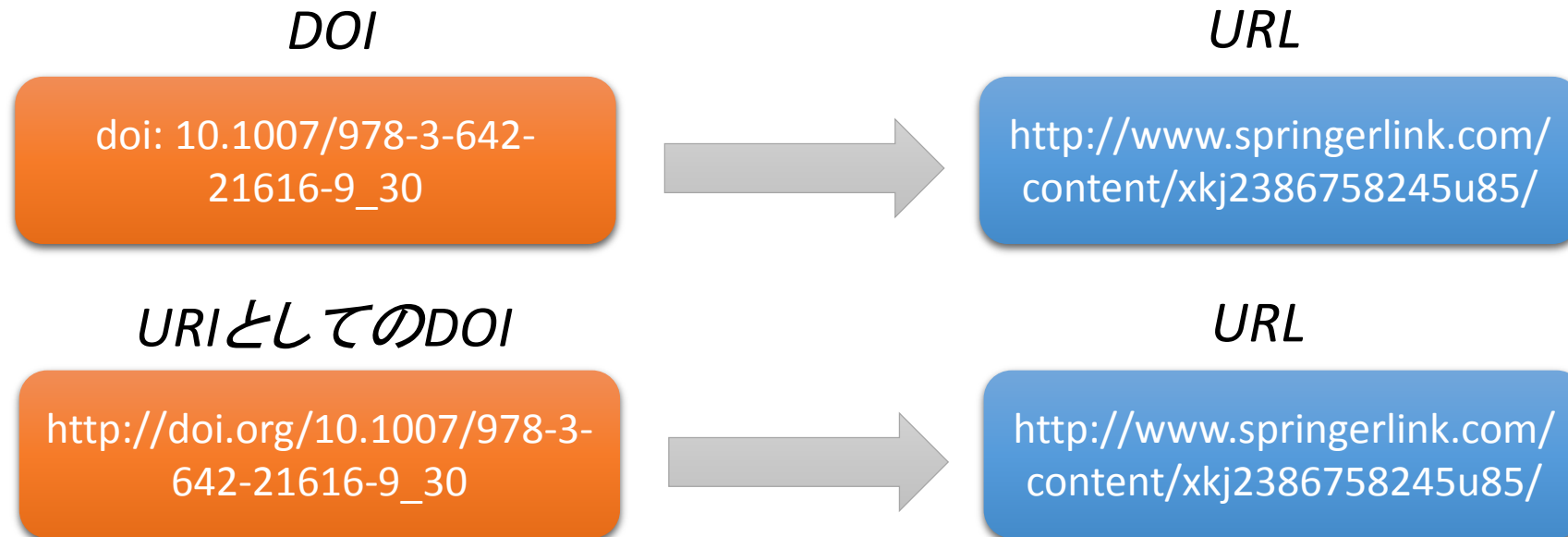
# DOI (Digital Object Identifier)

- 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 (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



# 基本的なDOI/CrossRefの使い方

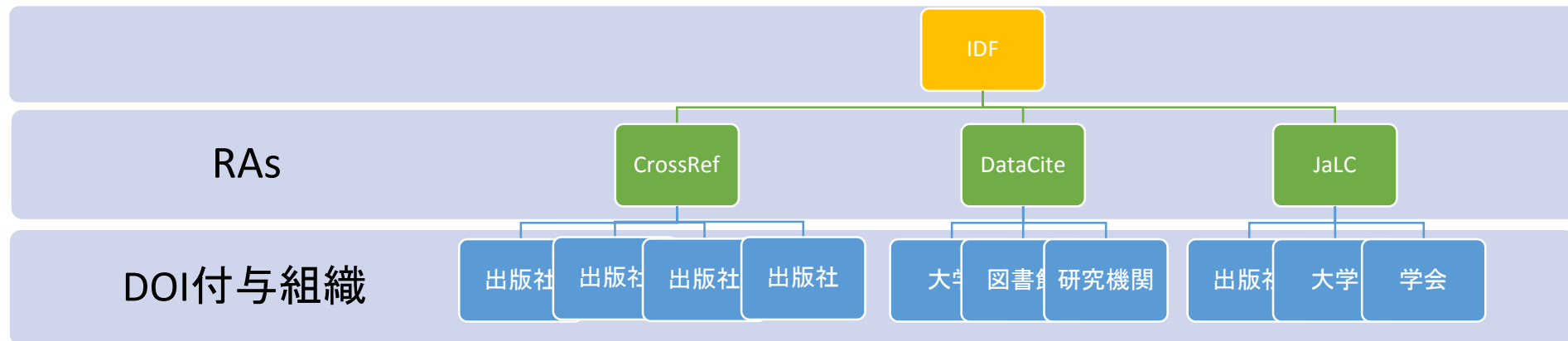
- 論文にDOIを振って論文公開(出版社)
- 引用文献にもDOIを振って論文公開(出版社)[CrossRef]
- 利用時・引用時にDOIを付記する(論文利用者・引用者)
  - Web環境下ならハイパーリンクとして
    - 例) <http://dx.doi.org/10.1000/182>
  - それ以外の環境ならテキストで
    - 例) DOI: 10.1000/182

# DOI/CrossRefのメリット

- 論文著者
  - 自論文が読まれやすく・引用されやすくなる
- 論文読者
  - 確実に論文の本文テキストにたどり着ける
  - 引用文献の本文テキストにたどり着ける[CrossRef]
- 出版社
  - 確実に自社サイトに誘導できる
  - 引用文献の同定が楽になる[CrossRef]
- 資金提供者、大学管理者
  - 業績確認のために論文の有無が簡単にわかる

# DOI運営の構造

- International DOI Foundation (IDF), 登録機関(Registration Agency, RA)、DOI付与組織、の3層構造
- RAはIDFに対してレジストリDB登録、レジストリDB運営、会費などで貢献
- RAは自身の顧客(会員、契約者)にDOI登録サービスの提供
- DOI付与組織は自分のコンテンツに対してDOIを発行し、自身の契約するRAを通じてDOI登録



# IDF Registration Agencies

## • CrossRef



- 学術的な専門的な研究コンテンツ
- 雑誌論文、図書、学会予稿など
- 引用リンク、検索可能なメタデータデータベース

## • mEDRA



- インターネット上の文書のための永続的引用システム
- 知的所有実体 (intellectual property entities) 相互の関係トラッキング
- タイムスタンプや電子署名の認定



## • OPOCE (Office des publications EU)

- EU 配下組織の出版物のためのDOI 管理を通じ、EU の公式パブリッシャーとして機能



## • China National Knowledge Infrastructure (CNKI)

- 中国の政治・経済・人文科学・科学技術などの情報。

## • DataCite



- データセットに関する学術基盤構築。ベストプラクティスの共有、
- 識別や名前解決の課題などへの取り組み

## • EIDR



- 映画、テレビ番組、商業音声／映像資料へのレジストリ

## • Wanfang Data Co., Ltd.

- 中国語雑誌のリンクサービス構築と、中国の科学データ管理

## • Airiti, Inc

- 繁体中国語資料へのDOI



## • ISTIC (The Institute of Scientific

and Technical Information of China)

- 中国語学術雑誌・中国の科学データ。



## • Japan Link Center (JaLC)



# CrossRef

- STM出版における引用関係を管理する仕組み
- 1999年～
- DOIのRA
- 世界70カ国に会員を持ち、現在登録されているDOIの大部分が CrossRefのDOI。
- 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

# 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

The screenshot shows the DataCite website homepage. At the top is a blue navigation bar with links for Home, Members, FAQs, Services, Resources, Events, and Contact us. Below the navigation bar is the DataCite logo, which consists of a stylized blue 'D' with a circular arrow around it, followed by the text 'DataCite' and the tagline 'Helping you to find, access, and reuse data'. The main content area features three article teasers, each with a 'Read more' link. The first article is 'DataCite Business Models Principles are online', published by Frauke Ziedorn on 31 October 2012. The second is 'DataCite Services', published on 18 October 2012. The third is 'Re-scheduled Open Hours for Datacenters', published on 10 October 2012. On the right side, there are three blue buttons with white text: 'Why cite data?', 'What is DataCite?', and 'What do we do?'. Below these buttons is a 'DOI resolver' section with a text input field and a 'Submit' button. At the bottom right, there is a 'Metadata Search' section with a search icon.



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](https://hdl.handle.net/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 Polarinstitut 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 \(NIPR\)](#) 🔍

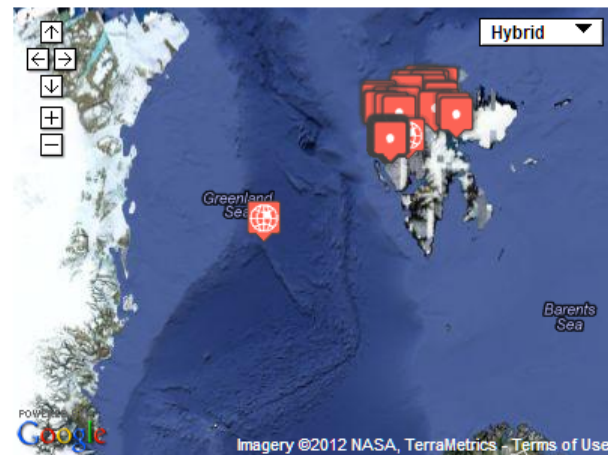
**Coverage:** *Median Latitude:* 79.212347 \* *Median Longitude:* 13.300383 \* *South-bound Latitude:* 74.150000 \* *West-bound Longitude:* -13.350000 \* *North-bound Latitude:* 80.183333 \* *East-bound Longitude:* 20.066667

*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

# Japan Link Center (JaLC)

- 平成24年3月に世界で第9番目に国際DOI財団(International DOI Foundation, IDF) から、DOI登録機関(RA)の認定を受けた。
- 同年5月に4国内学術機関による共同で運営を開始
  - 科学技術振興機構 (JST)
  - 国立情報学研究所 (NII)
  - 国立国会図書館 (NDL)
  - 物質・材料研究機構 (NIMS)

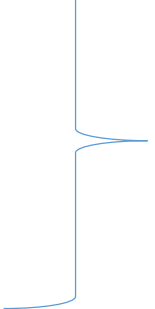
# ジャパン・リンク・センター(JaLC)設立趣旨

- JaLCはRAの一つとして、日本発の学術コンテンツ情報を収集し、普及、利用を促進する目的で2012年に設立された。
- 日本におけるRAとしてJaLCを設立した目的は、日本発の学術コンテンツの書誌情報を網羅的に収集することによって日本国内の利活用を促すと共に、世界から日本の研究成果へのアクセス環境を向上することにある。DOIは学術コミュニケーションの重要なツールとなり、国内の研究者がDOIを利用することは日常的に普及しつつも、国内でのDOI登録は浸透していない。主たる理由として、運営母体が海外の組織であるために、日本における学術コミュニケーションの事情に合わせた対応が困難であったことが挙げられる。日本の事情とは、具体的には日本語文献の管理、流通および、日本に存在する電子化された学術コンテンツの多様性、電子化由来の多様性などである。

# ジャパン・リンク・センター(JaLC)設立趣旨

- 日本における学術情報のセンターの必要性
  - 日本語文献の管理、流通
    - 日本語メタデータの適切な取り扱い(検索など)
    - 日本語文献への適切なメタデータ設計
  - 電子化された学術コンテンツの多様性
    - 大学紀要、多くの学会誌など
  - 電子化由来の多様性
    - J-Stage(JST)
    - ELS (NII)
    - 機関リポジトリ(各大学等)

# JaLC会員と対象コンテンツ

- 出版社
    - 学術出版(雑誌論文、書籍)
  - 学会
    - 学術出版(雑誌論文、書籍)
  - 大学、研究機関
    - 博士論文
    - 大学出版物(紀要、テクニカルレポートなど)
    - 機関リポジトリコンテンツ(教員出版物、OA公開物など)
    - 研究データ
  - 国立国会図書館
    - 自館出版物、パブリックドメイン出版物、博士論文
  - アグリゲータ:プラットフォーム提供者やハーベスターなど(JST、NIIなど)
    - 準会員の支援
- 
- CrossRefの主たる対象

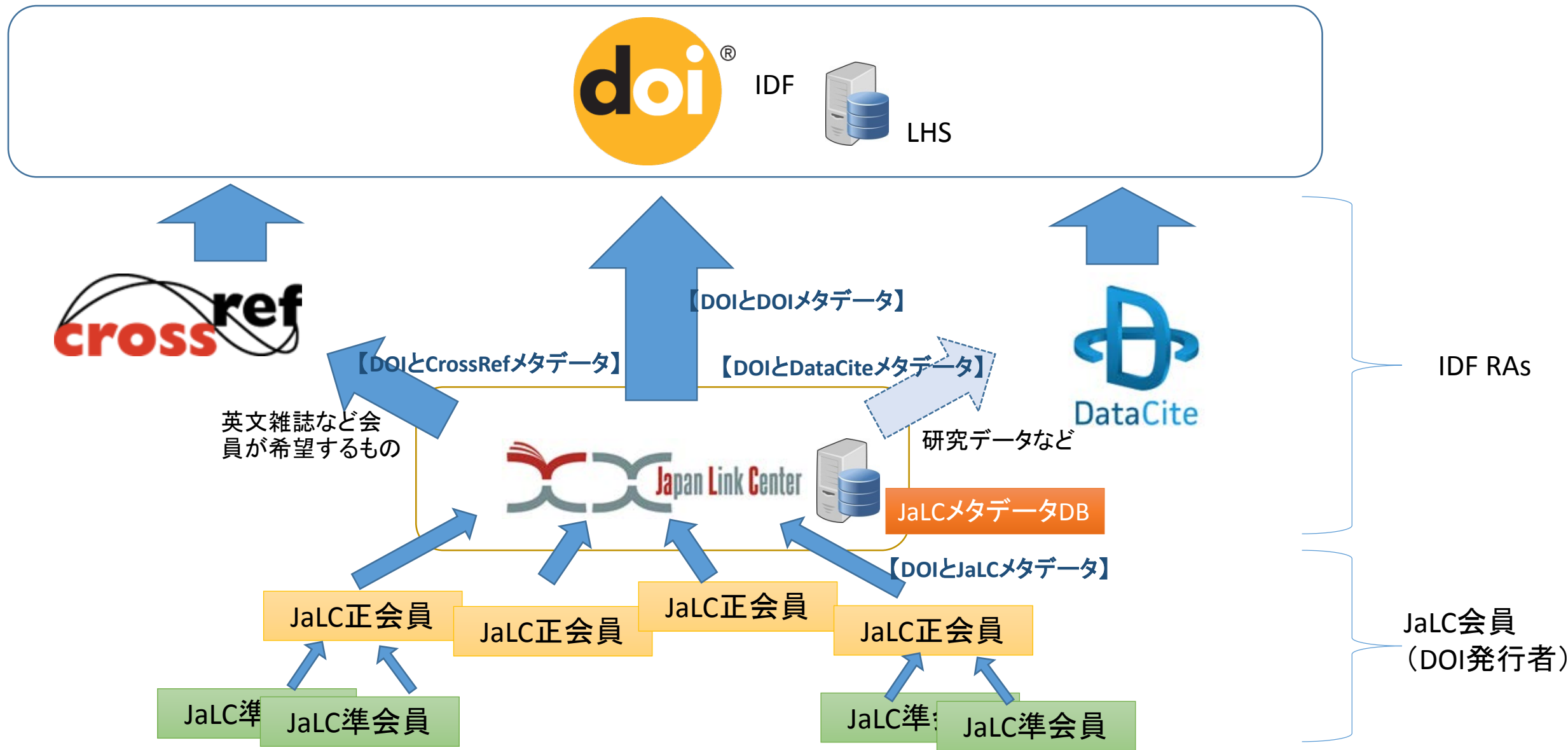
# JaLC DOIの目指す方向

- **研究者の業績が全てカバーできるDOI**
  - 多様な研究論文
    - 雑誌論文
    - 紀要論文
    - 書籍
    - 博士論文
    - 報告書
    - テクニカルレポート
    - . . . .
  - **データベース、データセット**
  - ソフトウェア
  - 教育用コンテンツ
  - . . . .

## メタデータ・スキーマ

ジャーナル・アーティクル  
書籍  
データ  
教育用コンテンツ  
汎用

# DOI登録の流れ



# 研究データへのDOI登録実験プロジェクト

- 2014年10月～2015年9月
- 目的
  - 日本における研究データへのDOI登録の仕組みを本プロジェクトへの参加機関とともに新たに構築する
  - 研究データへのDOIの持続性、対象の粒度、DOIの活用などを検討する
  - 運用方法の検討やDOI登録のテストを実施し、今後の安定運用につなげる
- ゴール
  - 研究データへのDOI登録のポリシーを策定する
  - 新JaLCシステムにおいて適切な運用フローを確立する。フロー確立に当たっては、DOI登録のテストを行って確認をとる



# 参加機関

機関	DOI登録対象とするデータの概要
科学技術振興機構	生命科学分野
国立極地研究所	地球科学・環境分野 (北極域、南極域における実験及び観測データ)
国立情報学研究所 (地球環境情報統融合プログラム (DIAS-P) )	地球科学データ (地上観測データ, 衛星観測データ、 気象予測モデル, 気候変動予測モデル、その他社会データ)
国立情報学研究所	機関リポジトリの保有する研究データ等 (テキスト、画像、音声)
産業技術総合研究所	地質および材料系
情報通信研究機構	超高層物理学分野 (地磁気データ、電離層データ、オーロラ画像 データ、衛星データ)
千葉大学附属図書館	一般的な研究データ、外部資金の助成による研究成果、 観測データ等のアーカイブ
物質・材料研究機構	<計画変更により再検討中>
理化学研究所 脳科学総合研究センター神経情報基盤センター	脳・神経科学分野

# 主なメタデータ項目

項目	参考:項目(英語)	説明
DOI	DOI	DOI
URL	URL	URL
タイトル	Title	データのタイトル
主題	Subject	主題、キーワード、分類、等
作成者	Creator	データ作成者の名前
所属機関	Affiliation	データ作成者の所属機関
研究者ID	Researcher ID	ORCID、e-Rad、科研費、等の番号
出版年月日	Publication date	研究データの出版・発行年
出版者	Publisher	研究データの出版者(研究機関等)
寄与者	Contributor	データ管理者、プロジェクトマネージャ、研究資金配分機関、等
版	Edition	Variation(出版版、著者版、等) Version(1.0、2.1等) Format(ファイルフォーマット)

項目	参考:項目(英語)	説明
関連コンテンツ	Related content	DOI、URL、relation type(引用、補足、別バージョンの存在、等)
代替識別子	Alternate ID	ローカルDBのIDなど
言語	Content language	コンテンツの言語
日付	Date	受理日、収集日、更新日、等
リソース種別	Resource type	データセット、ソフトウェア、等
権利	Rights	Creative Commons 等
追加情報	Description	抄録、研究方法、シリーズ情報、等
署名	Signature	データの品質保証等(JaLC独自)
位置情報	Geolocation	地域、場所の名前、緯度経度
研究資金情報	Fund	研究資金提供機関、課題番号、等
マルチプルレゾリューション 優先度	Multiple resolution priority	ミラーサイトやアーカイブサイトがある場合のランディングページ優先度

# 検討中の課題

## a) 運用フロー

- 取りまとめ機関の位置付け
- データのライフサイクルと担当者・担当機関
- DOI登録の対象データ
- Prefixの割り当て方針

## b) アクセスの持続性の保証

## c) DOI登録対象の粒度

## d) DOIのランディングページの要件の検討

## e) 研究データの特性に応じた取り扱い

- a) DOI登録後データ追加がある場合
- b) データ量が多い場合

## f) 研究データに登録したDOIの活用

# データのライフサイクルと担当者・担当機関

識別子(ID)

登録



メタデータ

作成



登録



修正



コンテンツ

作成



保存



公開



修正



破棄

# データのライフサイクルと担当者・担当機関

これまでの機関リポジトリ

図書館

識別子(ID)

登録

メタデータ

作成

登録

修正

研究者

コンテンツ

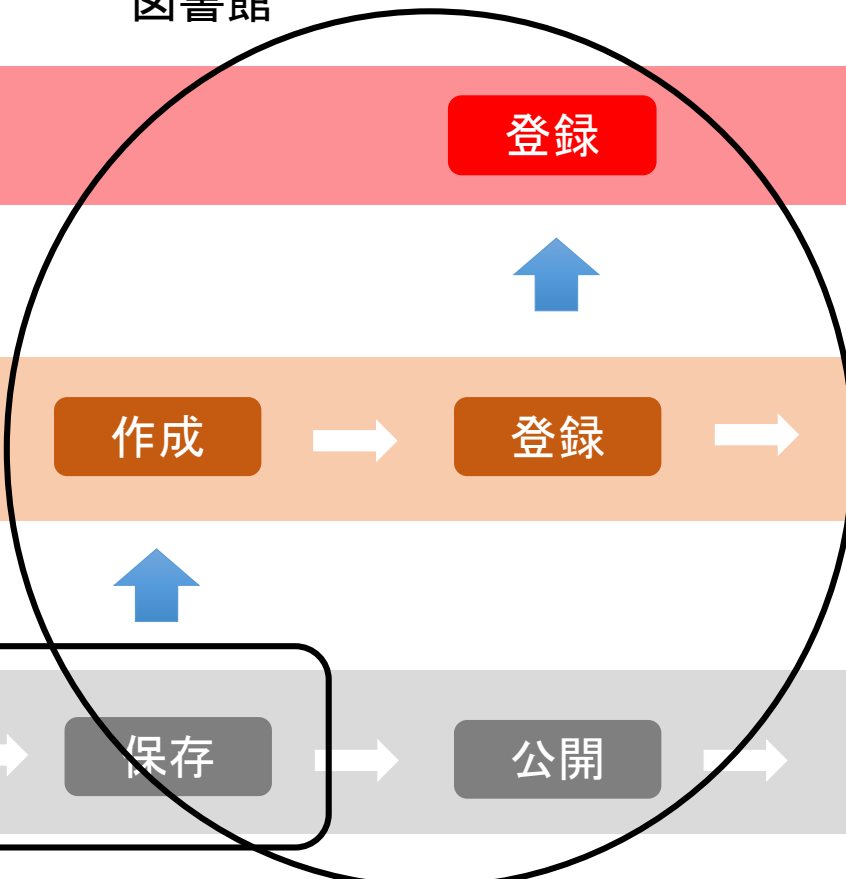
作成

保存

公開

修正

破棄



# データのライフサイクルと担当者・担当機関

データリポジトリ

図書館 研究機関

識別子(ID)

登録

メタデータ

プロジェクト

作成

登録

修正

研究者

コンテンツ

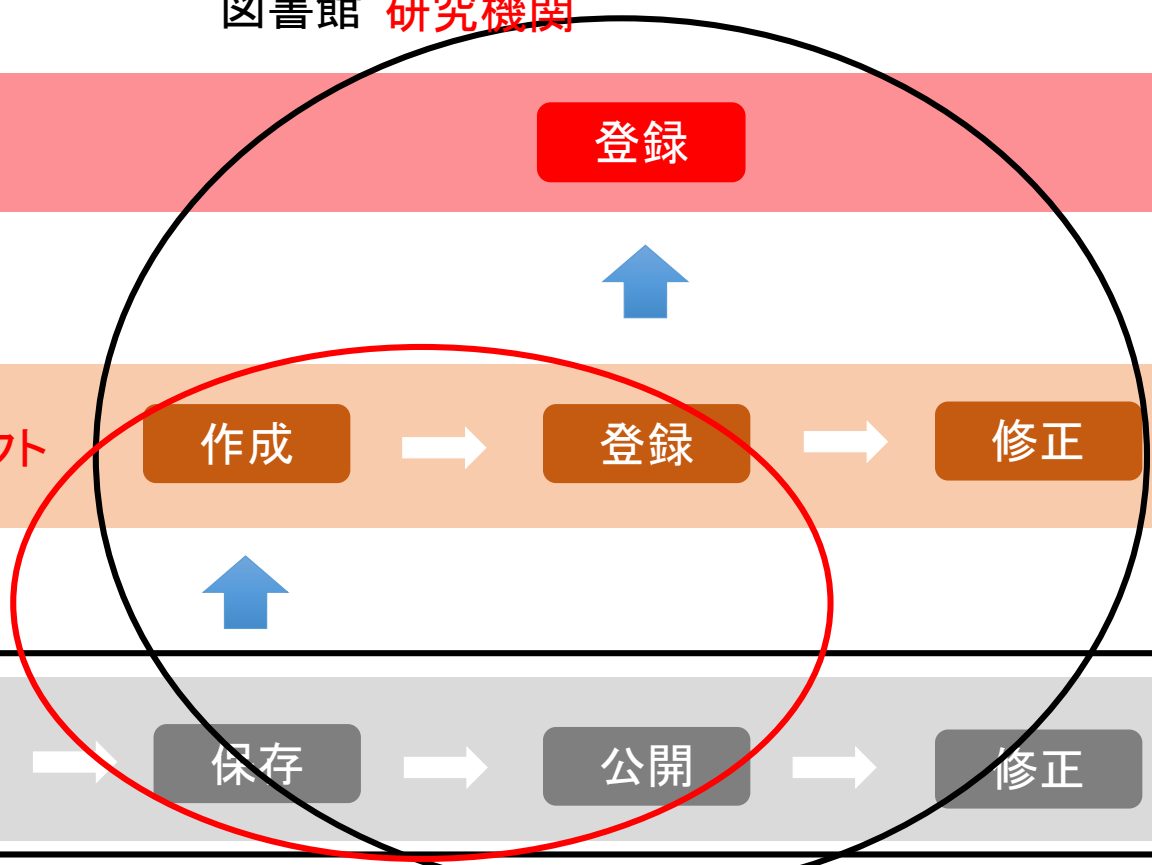
作成

保存

公開

修正

破棄



# 研究者の識別子

- 人の識別子  $\subset$  / デジタルオブジェクトの識別子
  - 人  $\neq$  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のメリット

## 大学

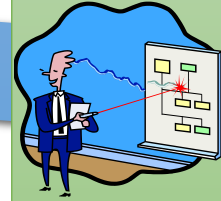


我が組織(学科、個別の研究者)の昨年の研究業績は？

うちからはどんなオープンアクセス論文をだしているのか？

我々の強い研究分野、弱い研究分野は？

## 学会



うちの会員の昨年の研究業績は？

うちの会員はどんなプロジェクトや研究助成、賞をとっているのだろうか？

どのくらい会員同士の共同研究はあるのだろうか？



## 研究者

過去5年に私が出版したものは？

興味あるこの論文の著者の他の論文は？

**46533489** 研究協力者をみつけるには？

## 研究助成機関



助成金応募者の著作論文は？

助成の結果として出版された論文は？

研究プロジェクトによってどんなデータセットが公開された？

## 出版社



ジャーナル投稿システムで効率的に著者と査読者を補足したい

既出版論文を含めてもっと著者のことが知りたい

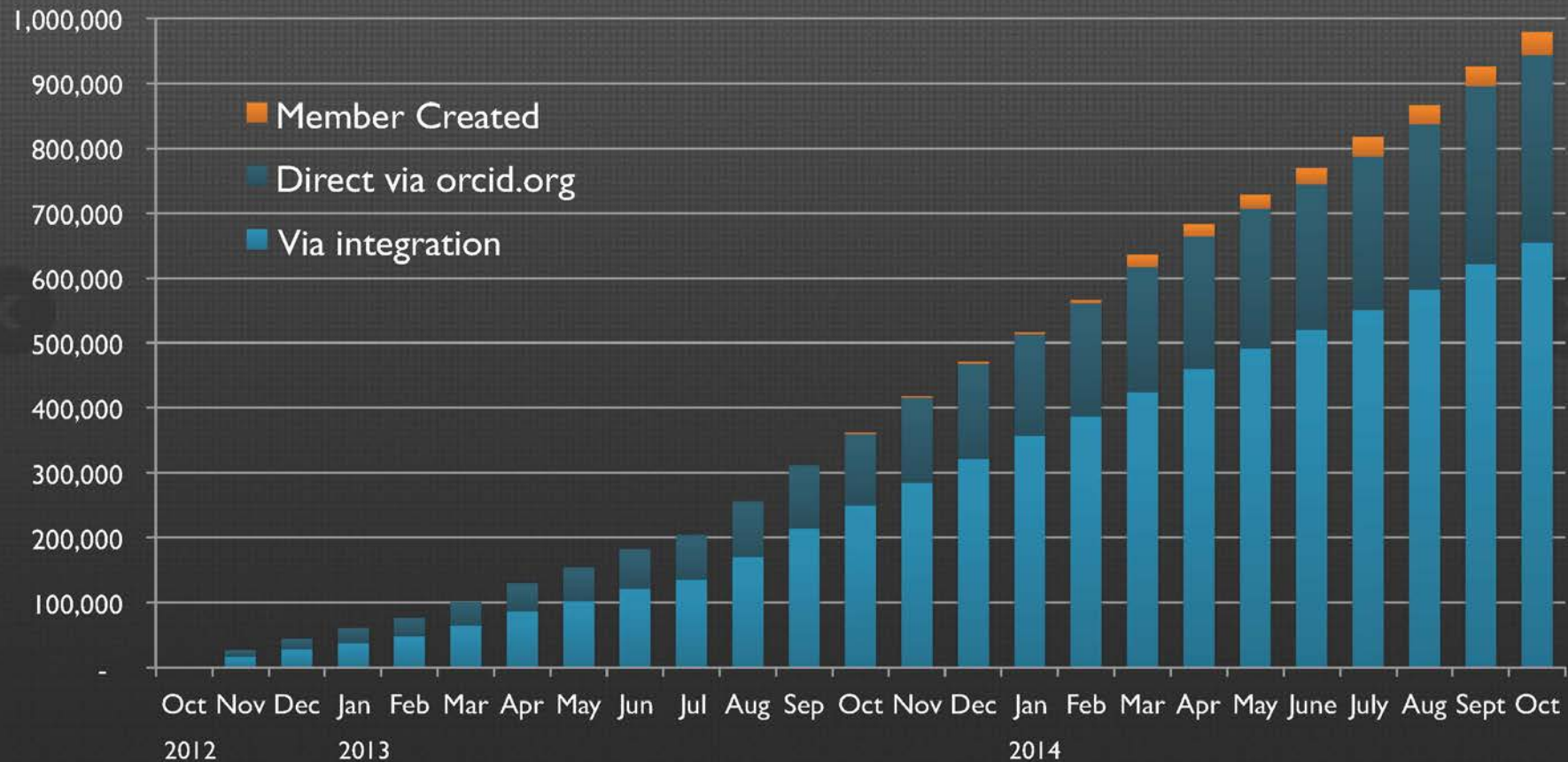
この著者と共同研究しうる査読者は？

ORCID enables machine-readable connections between iD and:

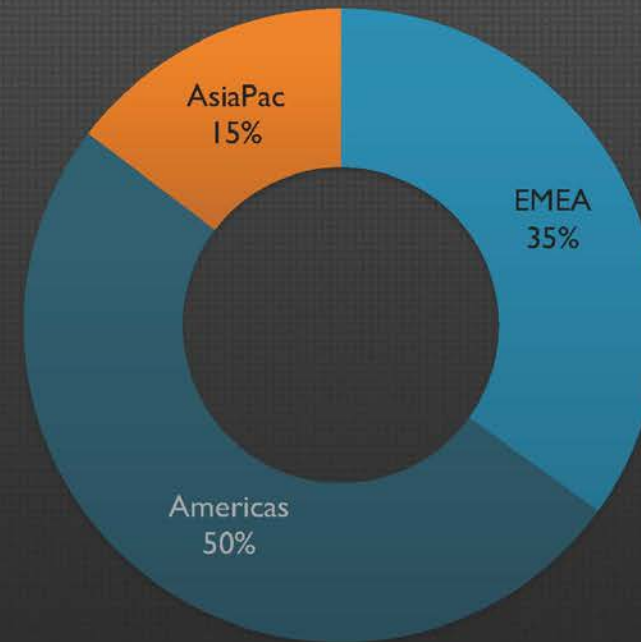
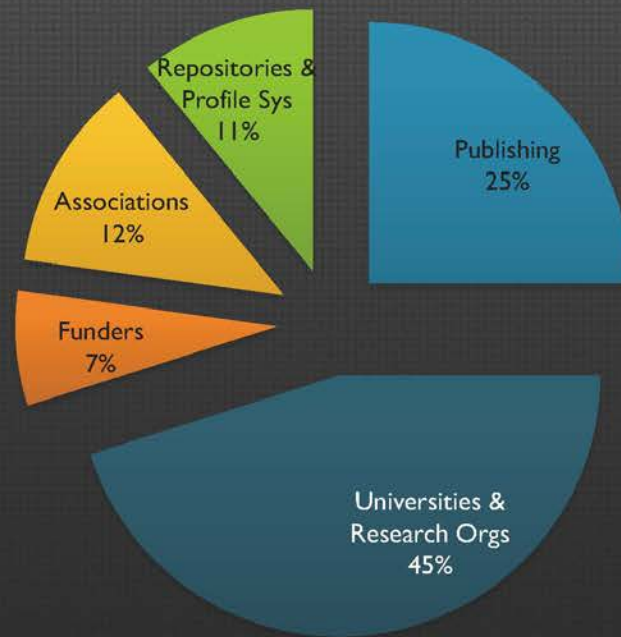
- works
- organizations
- other IDs



## ORCID nearing 1 million issued iDs



Over 160 members, from every region  
and sector of the international research  
community



# まとめ

- 研究成果はいずれ「データ」になる
- 研究データ流通基盤は必須の仕組み
- 研究データ流通はいくつかのレイヤー
  - 識別子、メタデータ・スキーマ、メタデータ、コンテンツ、フォーマット、リポジトリ
- DOIは研究データ流通の重要な識別子
  - これまでは論文に対する識別子
  - これからデータに対する識別子でもある。DataCite
  - 国内ではジャパン・リンク・センター(JaLC)がDOI登録機関
- JaLCでは現在、データDOI実験プロジェクト実施中
  - 国内では初めて分野を超えた研究データ関連機関の連携プロジェクト
- 沢山のTO DO
  - 世界における研究データ流通のそれぞれのレイヤーでの「協調と競争」にいかにかかわっていくか