

# 研究データTFの活動（全体報告）

国立情報学研究所 オープンサイエンス基盤研究センター  
尾城 孝一

# 今、なぜ研究データか？

1. オープンサイエンス推進の流れ
2. 研究公正（不正防止）の流れ

# オープンサイエンス推進

- G8科学大臣会合における研究データのオープン化を確約する共同声明  
(2013.6)  
→日本も調印(2013.6)
- 内閣府「国際的動向を踏まえたオープンサイエンスに関する検討会」報告書(2015.3)  
→公的研究資金による研究成果のうち、論文及び論文のエビデンスとしての研究データは原則公開  
→2015年度以降フォローアップ検討会を実施
- 第5期科学技術基本計画(2016.1)  
→知の基盤の強化に向けてオープンサイエンスを推進
- 科学技術・学術審議会 学術分科会 学術情報委員会「学術情報のオープン化の推進について(審議まとめ)」(2016.2)  
→研究資金配分機関、大学、NII、JST、学協会、国の役割の明確化
- 日本学術会議 オープンサイエンスの取組に関する検討委員会「オープンイノベーションに資するオープンサイエンスのあり方に関する提言」(2016.7)  
→研究データ基盤の整備、データ戦略、キャリアパス

# 研究公正（不正防止）

- 文部科学省
  - 『研究活動における不正行為への対応等に関するガイドライン』（2014.8.26）
  - 研究機関は、研究者に対して一定期間研究データを保存し、必要な場合に開示することを義務付ける旨の規程を設けるべし
- 日本学会協議
  - 『（回答）科学研究における健全性の向上について』（2015.3.6）
  - 論文等の形で発表された研究成果のもととなった実験データ等の研究資料は、当該論文等の発表から10年間の保存を原則とする
- 京都大学の事例
  - 『京都大学における公正な研究活動の推進等に関する規程』（2015.2.24）
  - 『京都大学の研究データ保存（理事裁定）』（2015.7.30）
  - 保存期間は、当該論文等の発表後少なくとも10年とし、これを下回って定めてはならない

# 大学図書館への期待

文科省・学術情報委員会「学術情報のオープン化の推進について（審議まとめ）」から抜粋

## 1. 研究データのオープン化

- 研究者のデータ管理計画作成の支援
- 研究データの保管に係る基盤整備に当たって、アカデミッククラウド（NIIが構築）の活用を図る

## 2. 研究成果の散逸等の防止

- 研究データの管理に係る規則を定め、散逸等を防止
- 研究データにDOIを付与

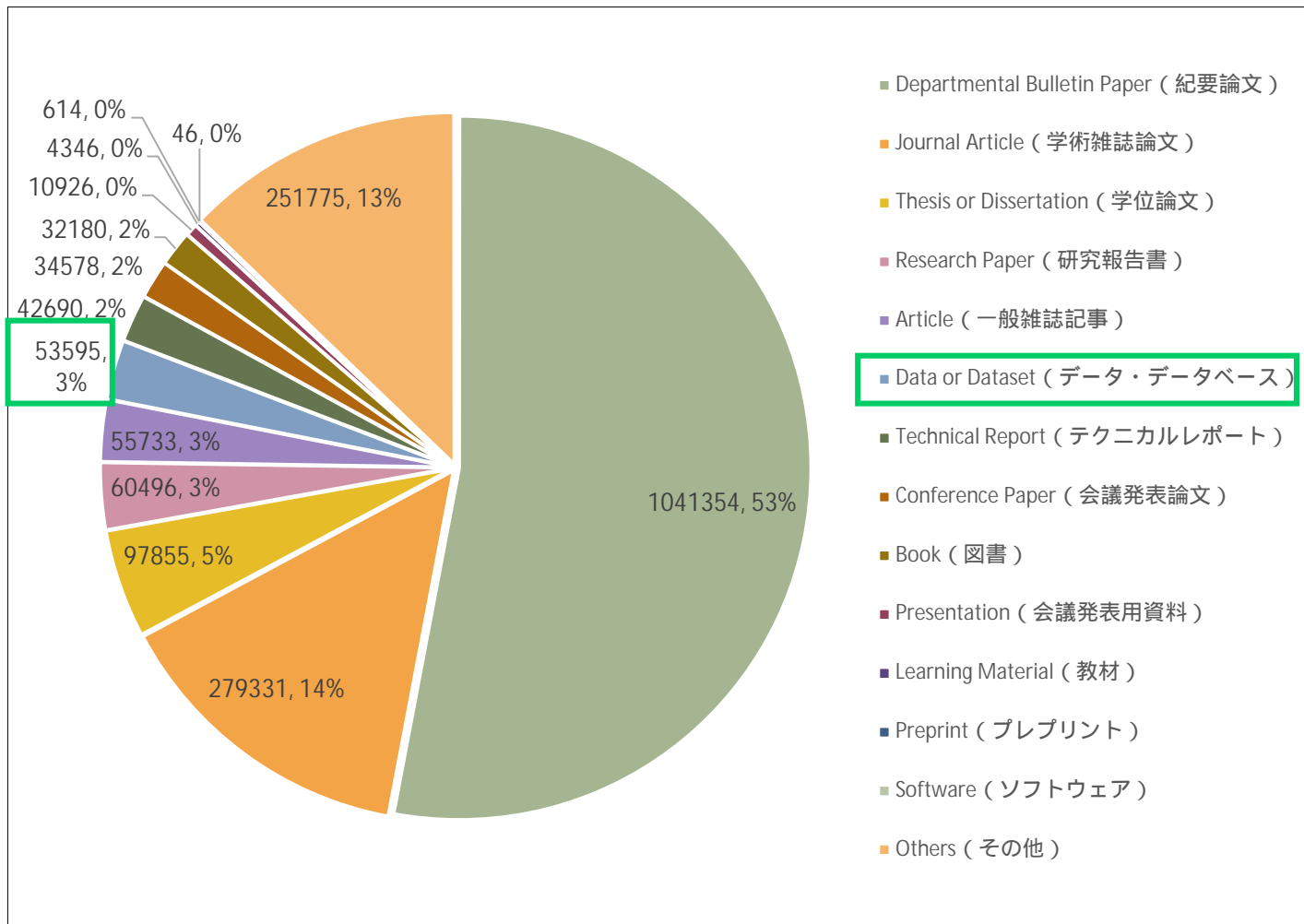
## 3. 人材

- 技術職員、URA及び大学図書館職員等を中心にデータ管理体制を構築
- 機関リポジトリの構築を進めてきた経験等から、研究成果の活用促進を担う

[http://www.mext.go.jp/b\\_menu/shingi/gijyutu/gijyutu4/036/houkoku/1368803.htm](http://www.mext.go.jp/b_menu/shingi/gijyutu/gijyutu4/036/houkoku/1368803.htm)

# 機関リポジトリのコンテンツ

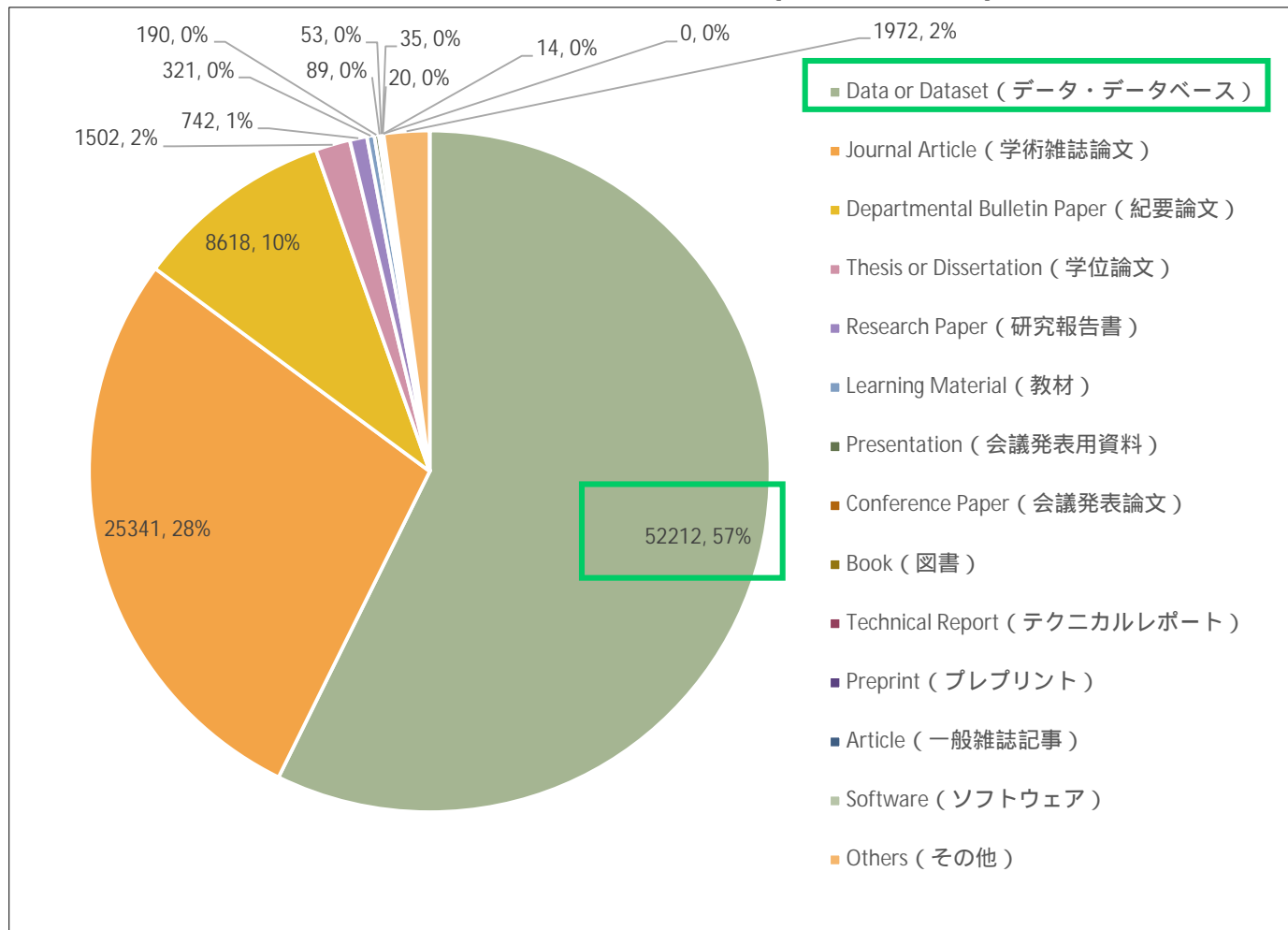
## 資源タイプ別コンテンツ（日本全体）



NII\_IRDBコンテンツ分析 2017.4 <http://irdb.nii.ac.jp/analysis/index.php>

# 機関リポジトリのコンテンツ

## 資源タイプ別コンテンツ (千葉大学)



NII\_IRDBコンテンツ分析 2017.4 <http://irdb.nii.ac.jp/analysis/index.php>

# 千葉大学学術成果リポジトリ (CURATOR)

CHIBA UNIVERSITY REPOSITORY  
for Access to Outcomes from Research

ログイン English

千葉大学学術成果リポジトリ ブラウズ ランキング

ブラウズ 資源タイプ

件数	資源タイプ
24	プレプリント Preprint
35	テクニカル・レポート Technical Report
83	会議発表論文 Conference Paper
8,646	紀要論文 Departmental Bulletin Paper
25,328	雑誌掲載論文 Journal Article
1,869	博士論文 Doctoral Thesis
744	研究報告書 Research Paper
5	単行書 Book
48	単行書の章 Book Chapter
321	教材 Teaching Material
51,818	データセット(萩庭) Dataset(Hagihiwa)
194	プレゼンテーション Presentation
13	一般雑誌記事 Article
1,992	その他 other

附属図書館 HP 千葉大学 HP

CURATOR Logo: Design

<http://opac.ll.chiba-u.jp/da/curator/?lang=0>



(萩庭さく葉データベース)



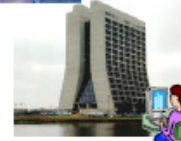
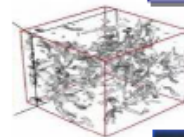
# 科学の方法論のパラダイムシフト

## 科学方法論のパラダイムシフト

- 数1000年前:
  - 実験科学 (Experimental Science)
    - 自然現象の記述
- 数100年前~
  - 理論科学 (Theoretical Science)
    - ニュートンの法則, マックスウェルの方程式...
- 数10年前~
  - 計算科学 (Computational Science)
    - 複雑な現象のシミュレーション
- 今日:
  - **eサイエンスあるいはデータ中心の科学**  
(e-Science or Data-centric Science)
    - 実験, 理論, シミュレーションの統合
    - データ探索やデータ・マイニングの活用
    - 「論文本文はもはやメタデータであり, 本当に大事なのは論文の裏にあるデータそのものである」(ある脳科学研究者)




$$\left(\frac{a}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$



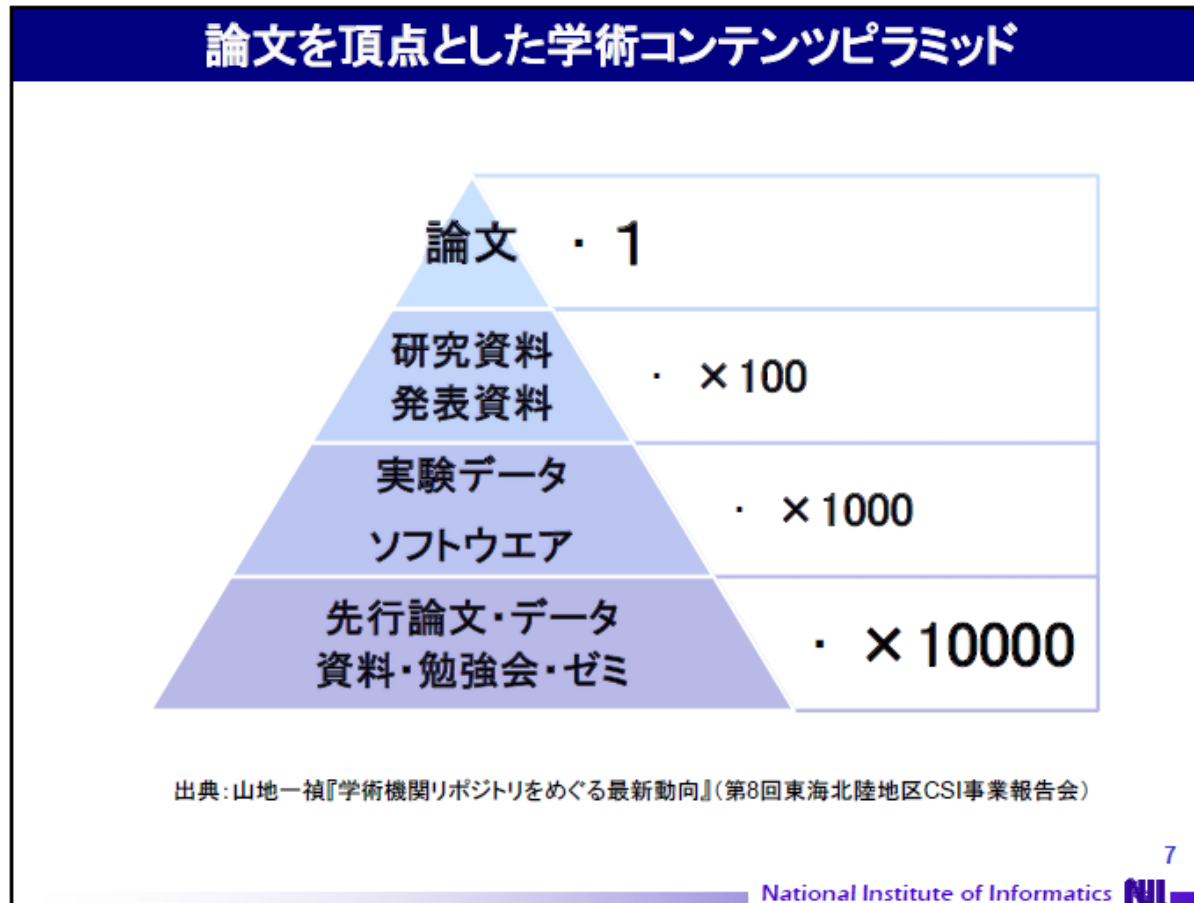
参考: Tony Hey, 2-3 November 2006, University of Stellenbosch, South Africa

5

National Institute of Informatics 

DRF/Share地域ワークショップ (北海道・東北地区)  
日時: 2008年12月11日(木) 13:00-17:00、12月12日(金) 9:00-16:00  
会場: 山形大学SCITAセンター

# コンテンツ・ピラミッド



DRF/Share地域ワークショップ (北海道・東北地区)  
日時: 2008年12月11日(木) 13:00-17:00、12月12日(金) 9:00-16:00  
会場: 山形大学SCITAセンター

# e-Scienceプロジェクト

CSI委託事業（2008～2009年）

「e-Science基盤構築のためのデータ・キュレーション機能拡充の実証実験」

この画像、あなたなら、どう検索しますか？



"Data-driven"

データの使い方を利用者に委ねることで研究の可能性を無限に広げてゆくのe-Scienceの魅力、でも、アクセスできなければ、その可能性は眠ったまま。

「最良の検索キーは、利用者が知っている」これが私たちの出発点です。

Folksonomy が、e-Science を変える。  
Social Tax が、データに命を与える。  
「みんな」で築く e-Science、リポジトリから始まります。

教員と図書館のコラボレーション

2008年、IRコンテンツにソーシャルタグを付与するシステムを開発。



「メタデータから漏れていたキーワードが、タグによって補われていた」  
「他のタグからの連想で、新たなタグが付与された」  
実証実験によって確認された効果の一例です。

2009年は「教員と図書館の連携」を固めながら、実証実験を拡大します。

e-Science基盤構築のためのデータ・キュレーション機能拡充の実証実験  
代表機関：千葉大学 / 連携機関：北海道大学、金沢大学、大阪大学、九州大学

テーマ： eサイエンスと機関リポジトリの連携の可能性についての調査・研究
プロジェクト名： e-Science 基盤構築のためのデータ・キュレーション機能拡充の実証実験 Webサイト
代表機関： 千葉大学
連携機関： 金沢大学 九州大学 北海道大学 大阪大学
実証実験として、コンテンツの作成者側と利用者側のそれぞれの立場から、e-Scienceデータ（主として千葉大学リモートセンシング研究センターの衛星画像）へ試行的にメタデータを付与できる環境を用意し、そのメタデータの評価をとおして、異種データの格納が可能なユーザコミュニティの需要に応えるメタデータ形式の検討を行い、人類の知的生産物の管理ツールとして機関リポジトリが位置づけられることをめざす。

千葉大学リモートセンシング研究センターの衛星画像のキュレーションに関する実証実験を期待したのだが、、、

結局は、植物標本の画像データに検索タグを付ける試みで終わってしまった

# 2016年度の活動とメンバー

- n RDMトレーニングツールの開発
- n ケーススタディによる研究データ管理ノウハウの蓄積

尾城 孝一

東京大学附属図書館 (主査)

山地 一禎

国立情報学研究所 (副主査)

前田 翔太

北海道大学附属図書館

三角 太郎

千葉大学附属図書館

天野 絵里子

京都大学学術研究支援室

大園 隼彦

岡山大学附属図書館

西園 由依

鹿児島大学学術情報部

南山 泰之

国立極地研究所情報図書室

# RDMトレーニングツールの開発

- 目的

- 学習者が研究データ管理（RDM: Research Data Management）に関する基礎的な知識を習得する
- RDMサービス構築の足掛かりを得る

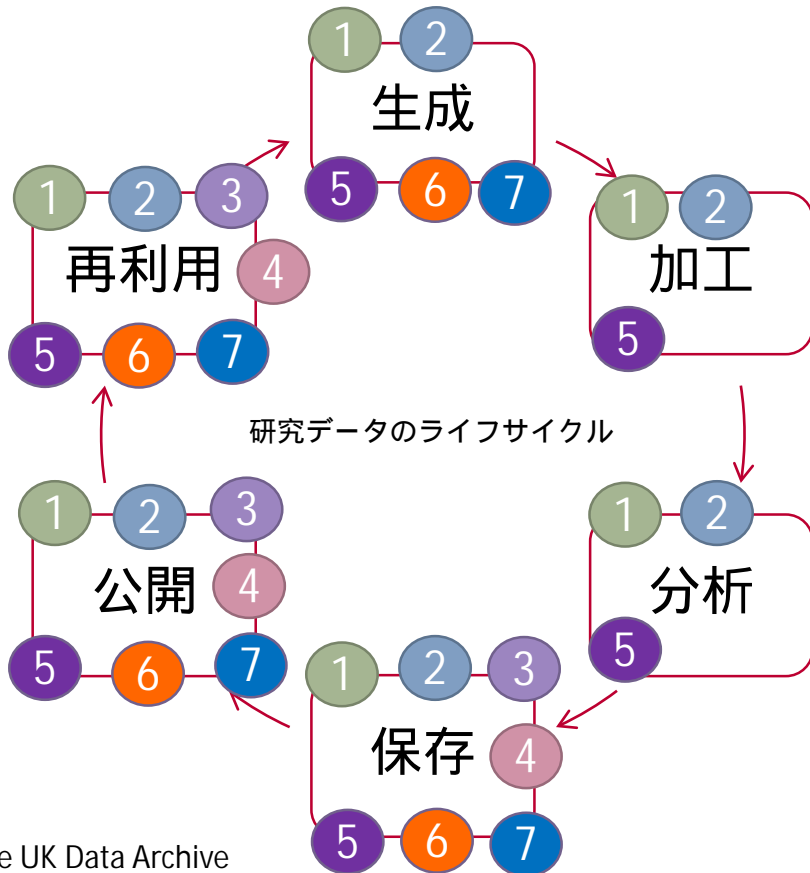
- 内容

- 全7章から構成されるeラーニング教材
- 各章は、スライドとスクリプトで提供し、確認テストやワークを取り入れる

- 受講対象者

- 図書館、IT部門、研究支援部門の職員（URA）等
- オープンサイエンスや研究データ管理に関心のある教職員

# データのライフサイクルと教材の構成



- 1章 導入
- 2章 データ管理計画
- 3章 保存と共有
- 4章 組織化・メタデータ
- 5章 法倫理的問題
- 6章 ポリシー
- 7章 サポートサービスの検討

• The UK Data Archive  
( <http://www.data-archive.ac.uk/create-manage/life-cycle> )  
• RDM Support - basic training course for information specialists  
( <http://dx.doi.org/10.6084/m9.figshare.1285313> ) より

# ケーススタディ

## 1. 消えゆくデータベースのレスキュー

- データベース科研やDnaviの調査
- 持続的な維持管理が困難な人社系データベースを機関リポジトリで救済する方策の検討

## 2. データジャーナルの刊行

- Polar Data Journalの創刊を通じて、データジャーナル刊行のためのプラットフォームの要件の検討
- 研究データの取扱いに関するノウハウの蓄積

# 消えゆくデータベースの兆候

- URLクリック → Not found
- 最終更新日が古い
- 運営組織、運営責任者が不明
- 検索できない、閲覧できない



# データベース・レスキュー

## 1. 対象DB予備調査

- 2015年度
  - 科研費研究成果公開促進費（H21-H23）で助成されたDB  
→文献のDB56件のうち約半数が適切に継続管理されていない
- 2016年度
  - Dnavi（国立国会図書館）を調査
    - 大学、研究所、高専等のDBは5,533件
    - 約半数が「文系」  
→その多くが適切に継続管理されていない可能性あり

## 2. メタデータ検討

- junii2スキーマでは研究データを保存するには不十分
- 2016年度にメタデータTFが開発したJPCOARスキーマで可能に！

# データ移行方式の検討

## 1. アイテム単位で登録

- 1件ずつ登録する、再利用可能なきれいなメタデータを付与できる  
工数がかかる

## 2. アイテム単位で一括登録

- DBと機関リポジトリのメタデータスキーマをマッピングし、一括登録  
マッピングが可能か

## 3. コレクション（DB）単位で登録

- データベースを階層化しコレクションを作り、コレクション毎に機関  
リポジトリに登録  
作業量は減少するが、再利用しにくい

## 4. 検索サイトを別途構築

- 機関リポジトリ側にはデータ + 既存のスキーマに合致するメタデータ  
のみを登録
- 別途、DB固有のメタデータを検索するためのサイトを構築  
検索サイトを構築するためのコスト

# データジャーナル

- データジャーナルとは
  - データそのものを記述する論文（データ論文）を掲載するジャーナル
    - 原著論文と同じ出版プロセス（査読・審査）
    - 「査読」によるデータの品質保証、再利用性の向上
- データジャーナルによって
  - データの利用者拡大
  - 検証可能性の向上
  - データ作成者、データ取得者の業績を明確に表現

# データ論文のサンプル

Earth System Science Data: The Data publishing Journal

Earth Syst. Sci. Data, 7, 157–171, 2015  
<http://www.earth-syst-sci-data.net/7/157/2015/>  
 doi:10.5194/essd-7-157-2015  
 © Author(s) 2015. This work is distributed under the Creative Commons Attribution 3.0 License.

Article Peer review Metrics Related articles

13 Jul 2015

**Filling the gaps in meteorological continuous data measured at FLUXNET sites with ERA-Interim reanalysis**

N. Vuichard<sup>1</sup> and D. Papale<sup>2,3</sup>  
<sup>1</sup>Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSU – UMR CE Merisiers, Bât 712, 91191 Gif-sur-Yvette, France  
<sup>2</sup>Department for Innovation in Biological, Agro-food and Forest systems (DIBAF),  
<sup>3</sup>Euro-Mediterranean Center on Climate Change (CMCC), Via Augusta Imperatore

Received: 28 Nov 2014 – Discussion started: 22 Jun 2015 – Accepted: 23 Jun 2015 – Published: 13 Jul 2015

**Abstract.** Exchanges of carbon, water and energy between the land surface and the atmosphere are measured by eddy covariance technique at the ecosystem level. Currently, the FLUXNET database contains more than 500 sites registered and up to 250 of them share data (free fair-use data set) for evaluating ecosystem models' performance, but this requires meteorological variables used as input. Because original in situ data often up to relatively long (some months) ones, we develop a new and robust method to downscale and correct the ERA-Interim data at site level. Our approach has the benefit of making use of (ERA-Interim) and a high temporal resolution spanning from 1989 to today site level, and for this reason a method to downscale and correct the ERA-Interim data to the level 4 data (L4) from the La Thuile collection, freely available. The performance of the developed method varies across sites and is also average over all sites, applying the bias correction method to the ERA-Interim data by 10 to 36 %, depending on the meteorological variable considered. In comparison to the internal variability of the in situ data, the root mean square error (RMSE) between the in situ data and the unbiased ERA-Interim data remains relatively large (on average over all sites, from 27 to 76 % of depending on the meteorological variable considered). The performance of speed field, in particular regarding its capacity to conserve a standard deviation similar to the one measured at FLUXNET stations.

<http://dx.doi.org/10.5194/ESSD.838234> sites can be downloaded

PANGAEA (地球環境分野のデータリポジトリ)

**PANGAEA**  
 Data Publisher for Earth & Environmental Sciences

SEARCH SUBMIT ABOUT CONTACT

**Citation:** Vuichard, Nicolas; Papale, Dario (2014): ERA interim reanalysis debiased at FLUXNET sites, doi:10.1594/PANGAEA.838234  
 Supplement to: Vuichard, N; Papale, D (2015): Filling the gaps in meteorological continuous data measured at FLUXNET sites with ERA-Interim reanalysis, Earth System Science Data, 7(2), 157–171, doi:10.5194/essd-7-157-2015

**Abstract:** Exchanges of carbon, water and energy between the land surface and the atmosphere are measured by eddy covariance technique at the ecosystem level. Currently, the FLUXNET database contains more than 500 sites registered and up to 250 of them share data (free fair-use data set) for evaluating ecosystem models' performance but it requires uninterpolated time series for the meteorological variables used as input. Because original in situ data often contain gaps, from very short (few hours) up to relatively long (some months), we develop a new and robust method for filling the gaps in meteorological data measured at site level. Our approach has the benefit of making use of continuous data available globally (ERA-Interim) and high temporal resolution spanning from 1989 to today. These data are however not measured at site level and for this reason a method to downscale and correct the ERA-Interim data is needed. We apply this method on the level 4 data (L4) from the La Thuile collection, freely available after registration under a fair-use policy. The performance of the developed method varies across sites and on average overall sites, the bias correction leads to correct from 10% to 36% of the initial bias between in situ and ERA-Interim data, depending of the meteorological variable considered. In comparison to the internal variability of the in situ data, the root mean square error (RMSE) between the in situ data and this unbiased ERA-Interim data remains relatively large (on average overall sites, from 27% to 76% of the standard deviation of in situ data, depending of the meteorological variable considered). The performance of the method remains low for the Wind Speed field, in particular regarding its capacity to conserve a standard deviation similar to the one measured at FLUXNET stations.

**Project data site ID:** 10.1594/PANGAEA.838234

**Location:** Median Longitude: 48.897560 • Median Longitude: -19.877200 • South-bound Latitude: -31.425000 • Shore-bound Longitude: 121.622000 • North-bound Latitude: 70.698000 • East-bound Longitude: 148.152000

**Abstract:** The dataset contains meteorological fields retrieved from ERA-Interim reanalysis and debiased making use of in situ data at FLUXNET stations. Data are interpolated from the original 3-hour time resolution of ERA-Interim product to the time resolution of measurements at FLUXNET stations (half-hour) and distributed under the same format than in situ data, in local time. Data are provided for the 123 sites from the La Thuile collection based on fair-use policy, as available in August 2013. This meteorological data set can be used to fill the gaps in original in situ data in order to get uninterpolated time series for the meteorological fields needed for running most of ecosystem models. Data are provided in csv files, sorted in folders by site and period. The meteorological fields contained in the dataset are the Air Temperature (Ta\_m, °C), the Vapour Pressure Deficit (VPD\_m, hPa), the Wind horizontal speed (WS\_m, m s<sup>-1</sup>), the Precipitation (Precip\_m, mm day<sup>-1</sup>), the Global Radiation (RG\_m, W m<sup>-2</sup>) and the Incoming Longwave Radiation (LWR\_m, W m<sup>-2</sup>).

Site ID	Name	Short Name	Site Principal Investigator	Method	Comment
1	Site 1	Site 1			
2	Site 2	Site 2			
3	Site 3	Site 3			
4	Site 4	Site 4			
5	Site 5	Site 5			
6	Site 6	Site 6			
7	Site 7	Site 7			
8	Site 8	Site 8			
9	Site 9	Site 9			
10	Site 10	Site 10			

**License:** CC BY 4.0 International  
 Creative Commons Attribution NonCommercial 3.0 Unported

**Size:** 612 data points

# Polar Data Journalの創刊

The screenshot shows the Polar Data Journal website. At the top, there is a blue header with the NIPR logo and the title 'Polar Data Journal'. Below the header, there is a 'Call For Paper' section with text explaining the journal's purpose and submission guidelines. To the right, there is a 'Submit Your Paper' button and a 'Menu' section with links for 'About This Journal', 'For Authors', 'For Reviewers', 'Policies', 'Contact Information', and 'History of Website'. Below the menu, there is an 'Announcements' section with a 'view number 10' link. At the bottom, there is a search bar and a 'Language' dropdown menu.

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

- 日本初のデータジャーナル（査読付き）
- 国立極地研究所が刊行
- JAIRO Cloudを利用
- JCには、データ論文（データの生成手法、特性、利用方法などのメタデータが蓄積される
- 実データは、外部データリポジトリに保存

南山泰之 [他] . 極域科学のデータジャーナル：Polar Data Journal創刊の取り組み.  
情報管理. 2017. 60(3). [掲載予定]

# プラットフォームの要件

- データリポジトリとの連携
  - バージョン管理
  - メタデータ上の連携
- データの保全
  - データの真正性・完全性の担保
  - 査読中のデータの保管と公開用のデータとの比較
- ジャーナルの流通基盤との連結
  - 分野横断の利活用を念頭におき、相互運用性を備えたプラットフォーム

# 今後の検討課題

- データリポジトリの信頼性
  - データの長期保存と、かつ改変されていないことを保証する仕組みが必要

参考：Core Trustworthy Data Repositories Requirements (RDA DSA-WDS partnership WG)

<https://doi.org/10.5281/zenodo.168411>

- データの相互運用性・再利用性
  - 標準的なデータフォーマット？
  - データのライセンスに関するガイドラインの必要性

参考：RDA & CODATA Legal Interoperability Of Research Data: Principles And Implementation Guidelines

<https://doi.org/10.5281/zenodo.162241>

- その他
  - データの「査読」
  - 差分データの取扱い

# 2017年度の活動予定とメンバー

## n RDMトレーニングツールの開発

- 教材の公開と普及
- JMOOC/gaccoでの開講（2017.11予定）
- 持続可能な教材開発体制の確立

## n データベースレスキュー

- データ移行の試行とワークフローの確立

## n データリポジトリ

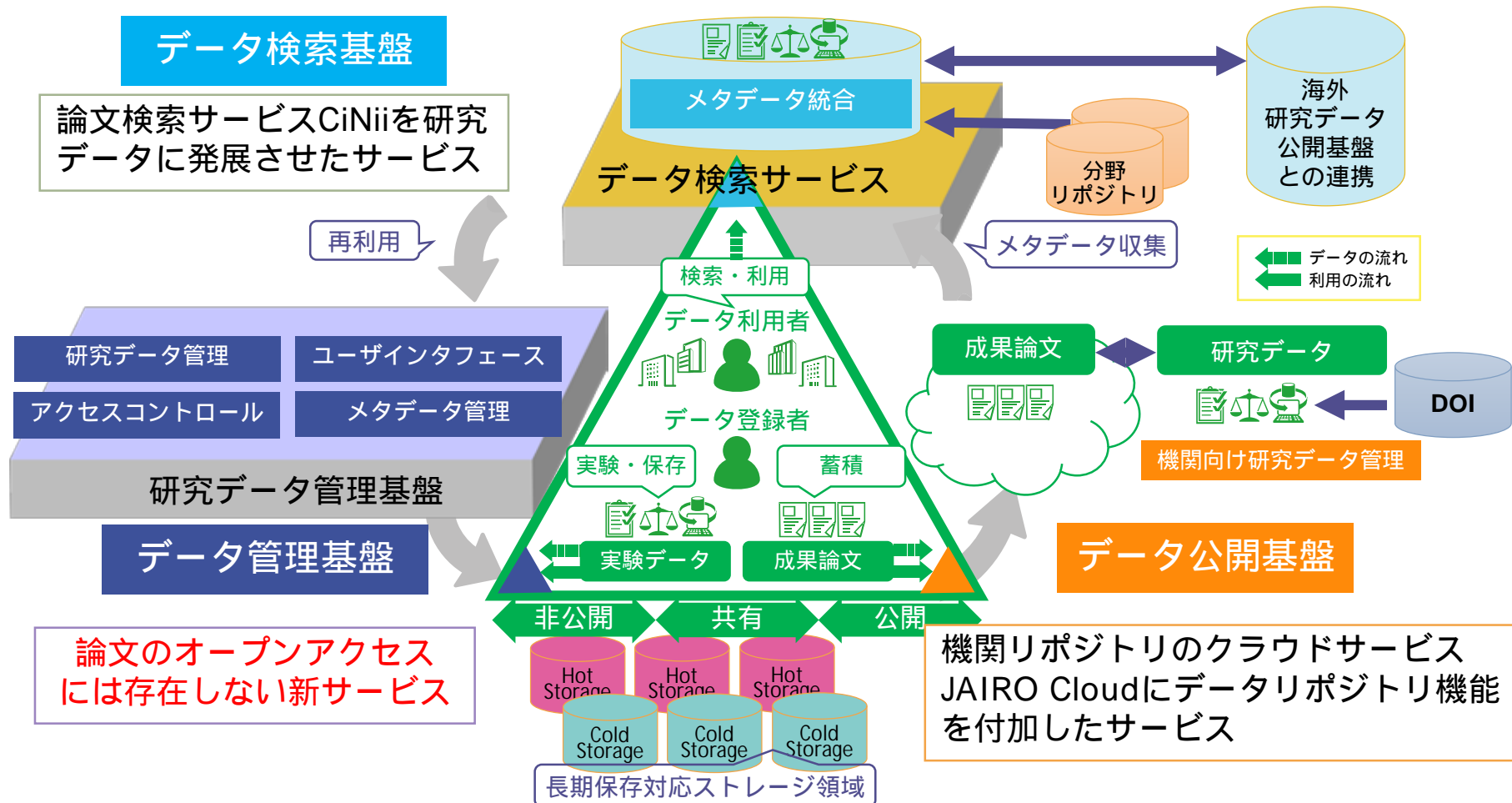
- データジャーナルの創刊から得た経験や知見をデータリポジトリの機能要件として整理
- JAIRO Cloudのデータリポジトリ化に活用

山地一禎	国立情報学研究所（主査）
常川真央	千葉大学附属図書館
松本侑子	東京大学附属図書館
青木学聡	京都大学情報環境機構
天野絵里子	京都大学学術研究支援室
西園由依	鹿児島大学学術情報部
南山泰之	国立極地研究所情報図書室
尾城孝一	国立情報学研究所



# 共に考え、共に創る研究データ基盤

これまでに研究データの共有や公開の機会がなかった分野の研究者も活用できる基盤として学際的なオープンサイエンスの実現を目指す



# 研究活動における研究データ基盤の役割

