2019年3月16日 (土) @京都大学
デジタルアーカイブ学会　第３回研究大会

学術論文出版からみる
研究データへのDOI付与とその粒度の考え方の課題

国立研究開発法人国立環境研究所
環境情報部
尾鷲瑞穂（Mizuho Owashi）

CC BY

National Institute for Environmental Studies, Japan

DOI(Digital Object Identifier)とは

・デジタルオブジェクト識別子DOI(Digital Object Identifier)は、学術論文等のデジタルコンテンツに付与される国際的な識別子。

・学術論文、図書、研究データ等にこのDOIを付与・登録し、コンテンツの所在情報(URL)とともに管理され、恒久的なアクセスを実現することを目的としている。

DOIの表示形式と役割

National Institute for Environmental Studies, Japan

・デジタルな研究資源に付与される識別子

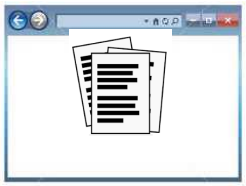
論文の識別子(固有の名前）としての役割

10.17595/20201222.001

プレフィックス　サフィックス

苗字　名前

・恒久的にデジタル資料の所在を一意に示す



論文の在り処を示す住所の役割

https://doi.org/10.17595/20201222.001

URI(Uniform Resource Identifier)

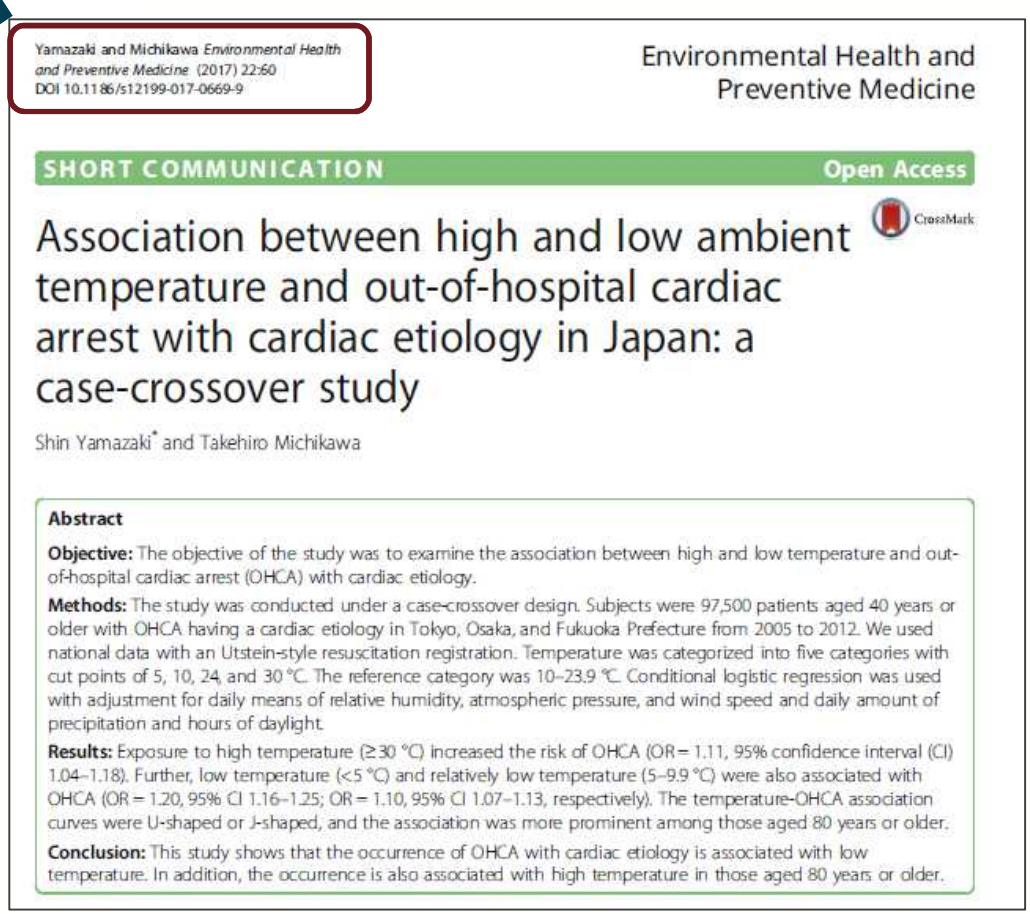
住所（戸籍）＝住む場所を変えても変わらない

論文に付与されたDOI

National Institute for Environmental Studies, Japan

Yamazaki and Michikawa *Environmental Health and Preventive Medicine* (2017) 22:60
DOI 10.1186/s12199-017-0669-9

Yamazaki S.; Michikawa T. Association between high and low ambient temperature and out-of-hospital cardiac arrest with cardiac etiology in Japan: a case-crossover study. Environmental Health and Preventive Medicine. 2017, 22:60.
DOI: 10.1186/s12199-017-0669-9



Yamazaki and Michikawa *Environmental Health and Preventive Medicine* (2017) 22:60
DOI 10.1186/s12199-017-0669-9

Environmental Health and Preventive Medicine

SHORT COMMUNICATION Open Access

Association between high and low ambient temperature and out-of-hospital cardiac arrest with cardiac etiology in Japan: a case-crossover study

Shin Yamazaki* and Takehiro Michikawa

Abstract

Objective: The objective of the study was to examine the association between high and low temperature and out-of-hospital cardiac arrest (OHCA) with cardiac etiology.

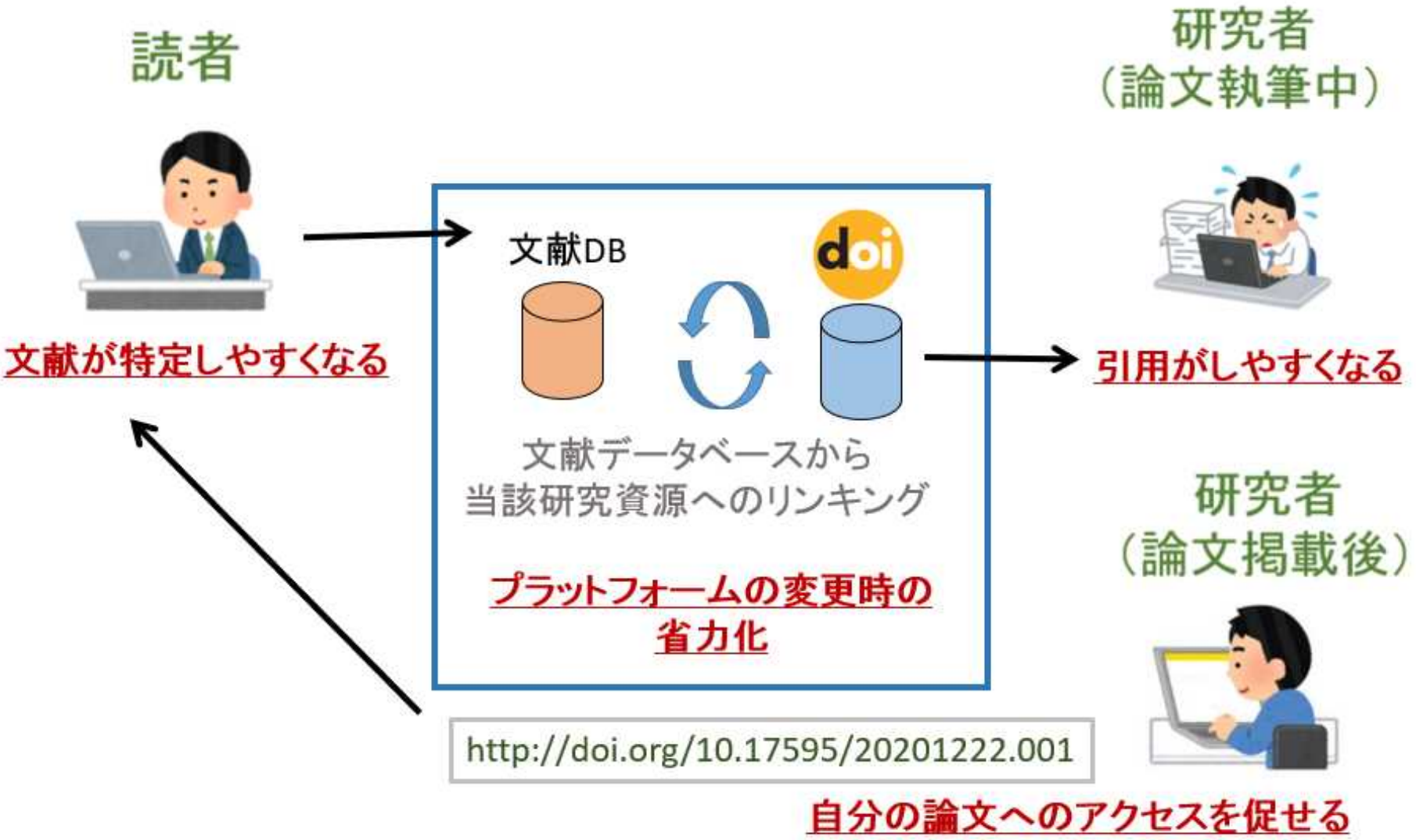
Methods: The study was conducted under a case-crossover design. Subjects were 97,500 patients aged 40 years or older with OHCA having a cardiac etiology in Tokyo, Osaka, and Fukuoka Prefecture from 2005 to 2012. We used national data with an Utstein-style resuscitation registration. Temperature was categorized into five categories with cut points of 5, 10, 24, and 30 °C. The reference category was 10–23.9 °C. Conditional logistic regression was used with adjustment for daily means of relative humidity, atmospheric pressure, and wind speed and daily amount of precipitation and hours of daylight.

Results: Exposure to high temperature (≥30 °C) increased the risk of OHCA (OR = 1.11, 95% confidence interval (CI) 1.04–1.18). Further, low temperature (<5 °C) and relatively low temperature (5–9.9 °C) were also associated with OHCA (OR = 1.20, 95% CI 1.16–1.25; OR = 1.10, 95% CI 1.07–1.13, respectively). The temperature-OHCA association curves were U-shaped or J-shaped, and the association was more prominent among those aged 80 years or older.

Conclusion: This study shows that the occurrence of OHCA with cardiac etiology is associated with low temperature. In addition, the occurrence is also associated with high temperature in those aged 80 years or older.

学術論文では普及が進んでいる

DOI付与のメリット（論文の場合）



DOIが付与されるコンテンツ

National Institute for Environmental Studies, Japan

ジャパンリンクセンター（JaLC）が想定しているコンテンツ

- ジャーナル/紀要/学術論文
- 図書/報告書
- e-learning/教材
- 研究データ　　など

長期的に運用されるリポジトリ等に収録されること前提
基本的に、デジタルなコンテンツであれば付与が可能

出版社の動向（PLOSの例）

National Institute for Environmental Studies, Japan

- 著者に対して研究上の発見に関わるデータを、論文公開後すぐに、制限なく利用できるようにすること。
- 論文の冒頭に、データがどこで、どのように利用可能かを示した”Data Availability Statement”を書くこと。

Data Availability

The following policy applies to all PLOS journals, unless otherwise noted.

PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction at the time of publication. When specific legal or ethical requirements prohibit public sharing of a dataset, authors must indicate how researchers may obtain access to the data.

When submitting a manuscript, authors must provide a *Data Availability Statement* describing compliance with PLOS's policy. If the article is accepted for publication, the data availability statement will be published as part of the accepted article.

Refusal to share data and related metadata and methods in accordance with this policy will be grounds for rejection. PLOS journal editors encourage researchers to contact them if they encounter difficulties in obtaining data from articles published in PLOS journals. If restrictions on access to data come to light after publication, we reserve the right to post a correction, to contact the authors' institutions and funders, or in extreme cases to retract the publication.

PLOS ONE “Data Availability”より
https://journals.plos.org/plosone/s/data-availability

日本で初めてDOIが付与されたデータとそれをエビデンスとした論文

National Institute for Environmental Studies, Japan

Mesospheric wind velocity data (30min. mean) observed with MF radar at Poker Flat, Alaska

Horizontal wind velocity in the altitude range of approx. 60-90 km is observed with Poker Flat MF (medium frequency) radar, using the radar wave at 2.43 MHz. The radar receives weak radio echo signals returned from the weakly ionized atmosphere (ionospheric D-region) at the target altitudes, to deduce horizontal air motions (Murayama, Y., K. Igarashi, D. D. Rice, B. J. Watkins, R. L. Collins, K. Mizutani, Y. Saito, and S. Kainuma, Medium Frequency Radars in Japan and Alaska for Upper Atmosphere Observations, IEICE Trans., E83-B, pp.1996-2003, 2000). Poker Flat MF radar has been constructed as part of Japan-US joint research program of Arctic middle & upper atmosphere ("Alaska Project") in collaboration between National Institute of Information and Communications Technology, Japan (formerly Communications Research Laboratory), and Geophysical Institute, University of Alaska Fairbanks.

Data Citation

Citation: Alaska Project of NICT (CRL)-GI/UAF, Mesospheric wind velocity data (30min. mean) observed with MF radar at Poker Flat, Alaska, doi:10.17591/55838dbd6c0ad

General Characteristics

Parameters: Mesospheric horizontal wind velocity
Processing level:
Temporal resolution: 30 minutes
Start date: 1998-10-16T01:45:00
Stop date: -PT1D

Observatory Location

Observatory: Poker Flat, Alaska
Latitude: 65.1°
Longitude: -147.5°
(Note: Latitudes and longitudes are expressed in decimal degrees. Eastern longitudes and Northern latitudes are positive.)

Citation & Contact Information

Acknowledgements: Alaska Project of NICT (CRL)-GI/UAF, Mesospheric wind velocity data (30min. mean) observed with MF radar at Poker Flat, Alaska
Contact Persons: Yasuhiro Murayama(orcid:0000-0003-1129-334X)

Links

Data Access: http://salm
DOI: 10.17591/5 Digital Obj
How to Cite Datasets: http://www Ball, A. an Guides, Ec

Provider Version

1.0

Update History

2019-01-2
2016-03-1
2015-06-1
2015-06-19T12:55:24+0900

エビデンスデータ
doi:10.17591/55838dbd6c0ad

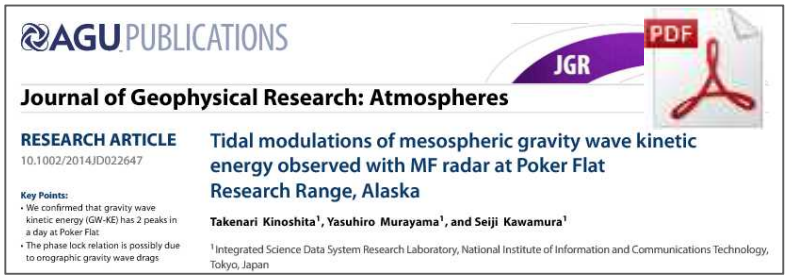
引用・被引用関係

論文
doi: 10.1002/2014JD022647

アラスカにおいて行われている上層大気の風速を電波観測したデータセット

Alaska Project of NICT (CRL)-GI/UAF, Mesospheric wind velocity data (30min. mean) observed with MF radar at Poker Flat, Alaska, doi:10.17591/55838dbd6c0ad

Kinoshita T, et.al.,Tidal modulations of mesospheric gravity wave kinetic energy observed with MF radar at Poker Flat Research Range, Alaska. JGR: Atmospheres. 2015, vol. 120, no. 13, p. 6379-6390. doi: 10.1002/2014JD022647



AGU PUBLICATIONS

JGR

PDF

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2014JD022647

Key Points:
- We confirmed that gravity wave kinetic energy (GW-KE) has 2 peaks in a day at Poker Flat
- The phase lock relation is possibly due to orographic gravity wave drags

Tidal modulations of mesospheric gravity wave kinetic energy observed with MF radar at Poker Flat Research Range, Alaska

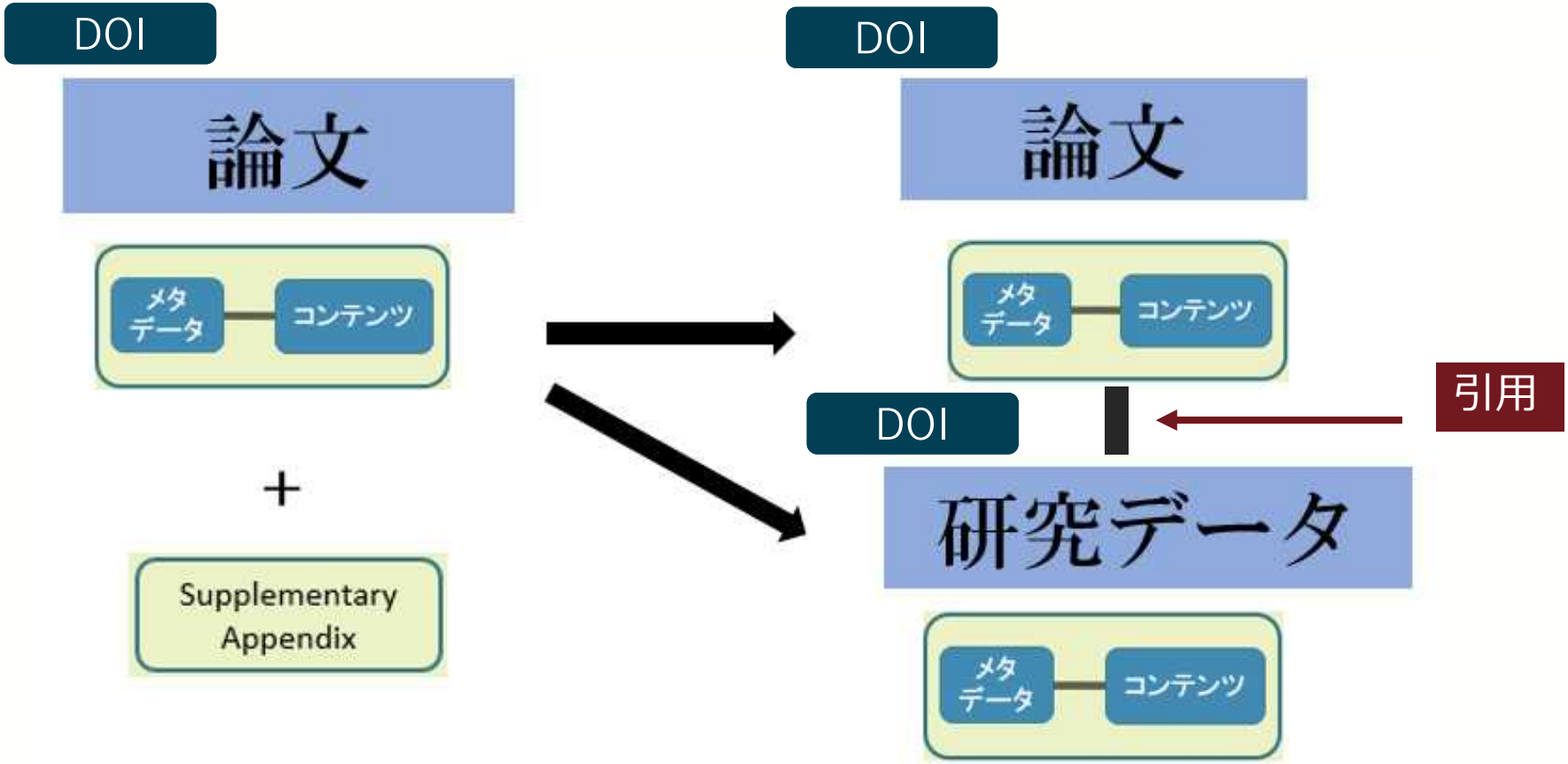
Takenari Kinoshita[1], Yasuhiro Murayama[1], and Seiji Kawamura[1]

[1]Integrated Science Data System Research Laboratory, National Institute of Information and Communications Technology, Tokyo, Japan

Acknowledgments

MF radar was manufactured by ATRAD Pty, Australia. The data of MF radar is available at SALMON (System for Arctic middle/upper atmosphere Observation data network) of National Institute of Information and Communications Technology, Japan. The dataset can be referred as doi:10.17591/55838dbd6c0ad. MERRA data are available from NASA. We thank Kaoru Sato for her fruitful discussions. We also thank Ulrike Langematz and anonymous reviewers for providing constructive comments. The GFD-DENNOU library was used for drawing figures.

論文と研究データの構造転換



研究データは論文とは個別に出版し、論文で引用してリンクさせる

データリポジトリPANGAEAのメタデータ

National Institute for Environmental Studies, Japan

研究データのDOI

PANGAEA.

Data Publisher for Earth & Environm ence

Download Data

Download **ZIP** file containing all datasets as tab-delimited text

Feakins, Sarah J; Eglinton, Timothy Ian; de Menocal, Peter B (2007): Biomarker records of DSDP Hole 24-231 and terrestrial sediments. *PANGAEA*, https://doi.org/10.1594/PANGAEA.763246,

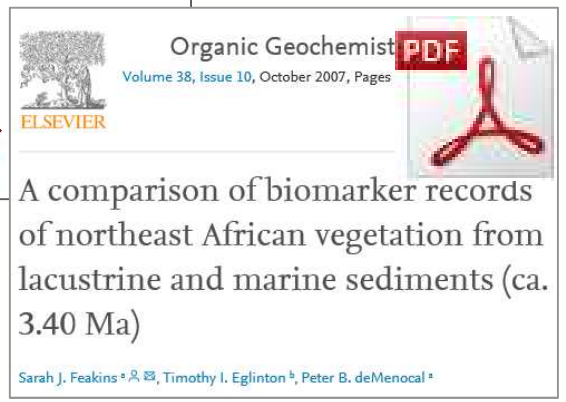
Supplement to: Feakins, SJ et al. (2007): A comparison of biomarker records of northeast African vegetation from lacustrine and marine sediments (ca. 3.40 Ma). *Organic Geochemistry*, **38(10)**, 1607-1624, https://doi.org/10.1016/j.orggeochem.2007.06.008

Always quote above citation when using data! You can do load the citation in several formats below.

RIS Citation | BibTeX Citation | Text Citation | Facebook | Tw | Google+ | Show Map | Google Earth

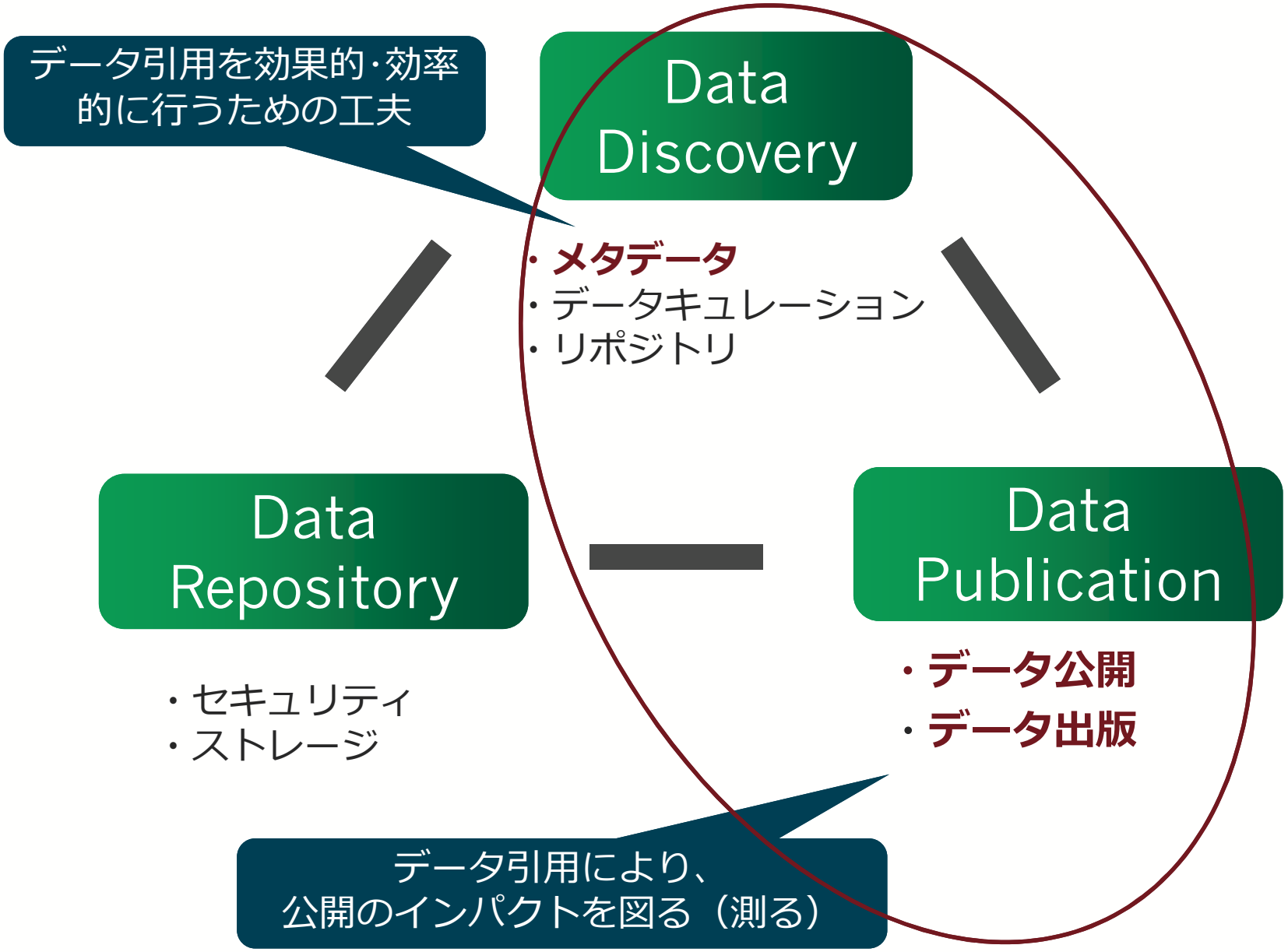
https://doi.pangaea.de/10.1594/PANGAEA.763246

論文のDOI



https://www.sciencedirect.com/science/article/pii/S0146638007001507

DOIが担う研究データインフラの役割



データジャーナルが次々と刊行

National Institute for Environmental Studies, Japan

SCIENTIFIC DATA

Nature Publishing Group

http://www.nature.com/sdata/

John Wiley & Sons

https://rmets.onlinelibrary.wiley.com/journal/20496060

Data in Brief

Elsevier B.V.

https://www.journals.elsevier.com/data-in-brief

Copernicus

https://www.earth-system-science-data.net/

Biodiversity Data Journal

Pensoft

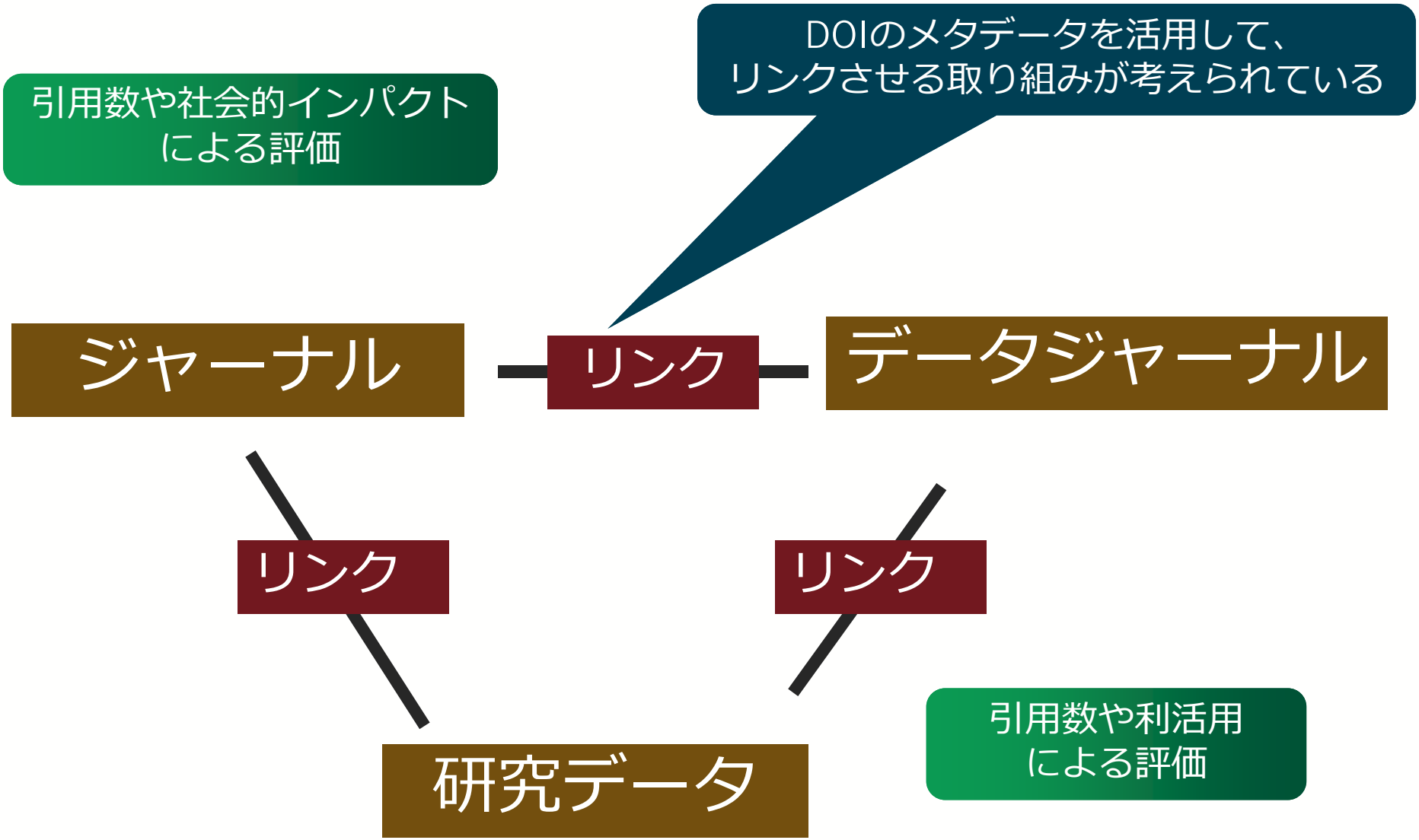
https://bdj.pensoft.net/

Polar Data Journal

NiPR（日本）

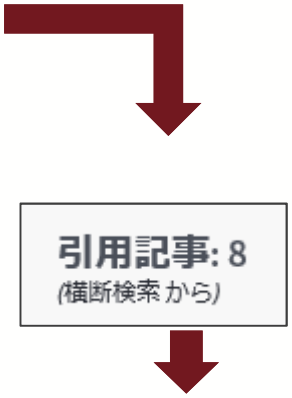
https://pdr.repo.nii.ac.jp/

データと論文を個別の成果として考える



Data Citation Index(WoS)の登場

研究データを論文記事と同じように検索。
引用された論文数が表示される。



Data Citation Indexの登場で、
論文の引用と同じようにデータの
引用数を測れるようになった。

【問題提起】

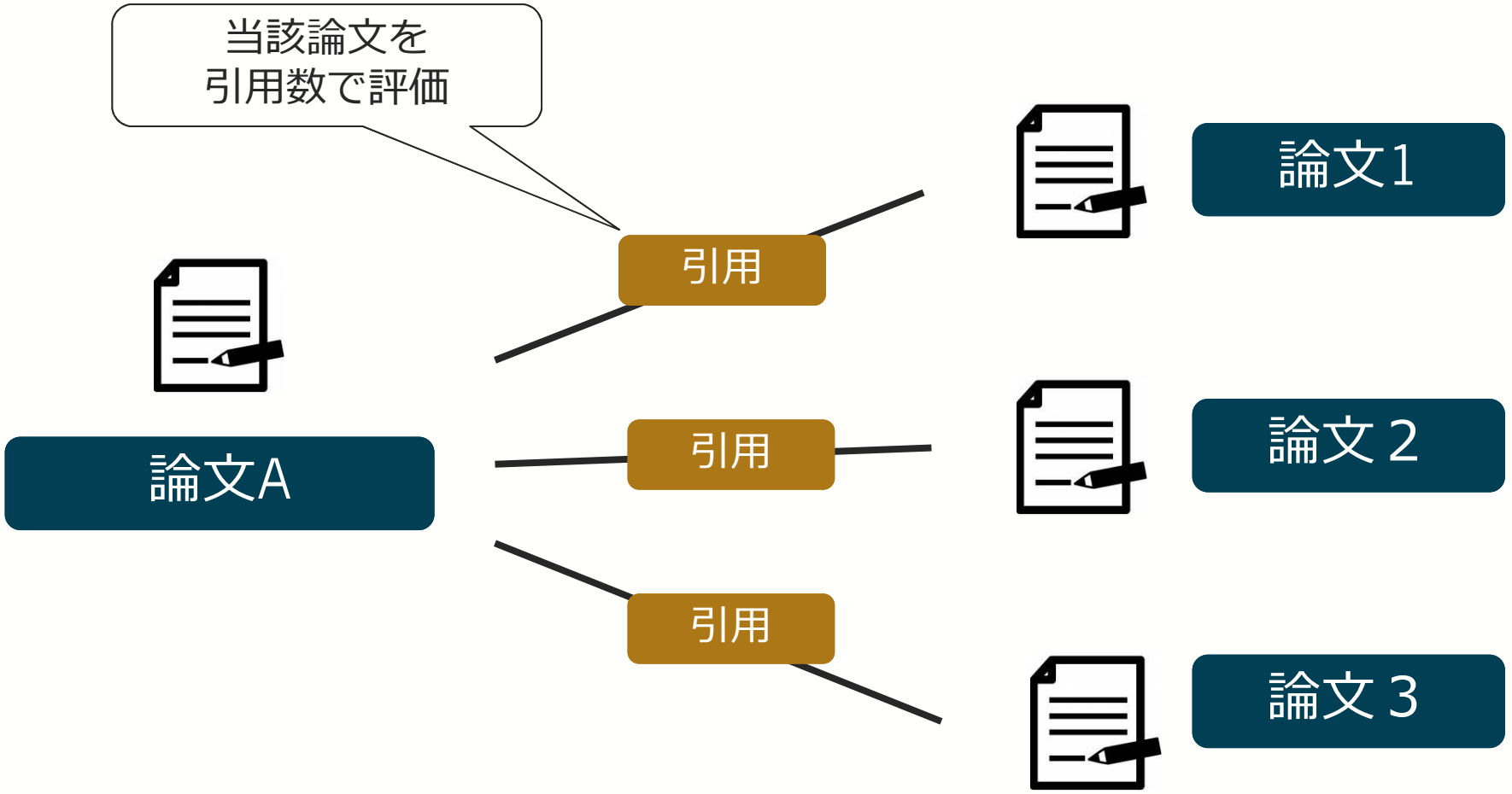
National Institute for Environmental Studies, Japan

・研究データへのDOIの付与により、論文と研究データの出版構造が変化した

・研究データも論文と同様に引用によって評価出来るようになった（なりつつある）

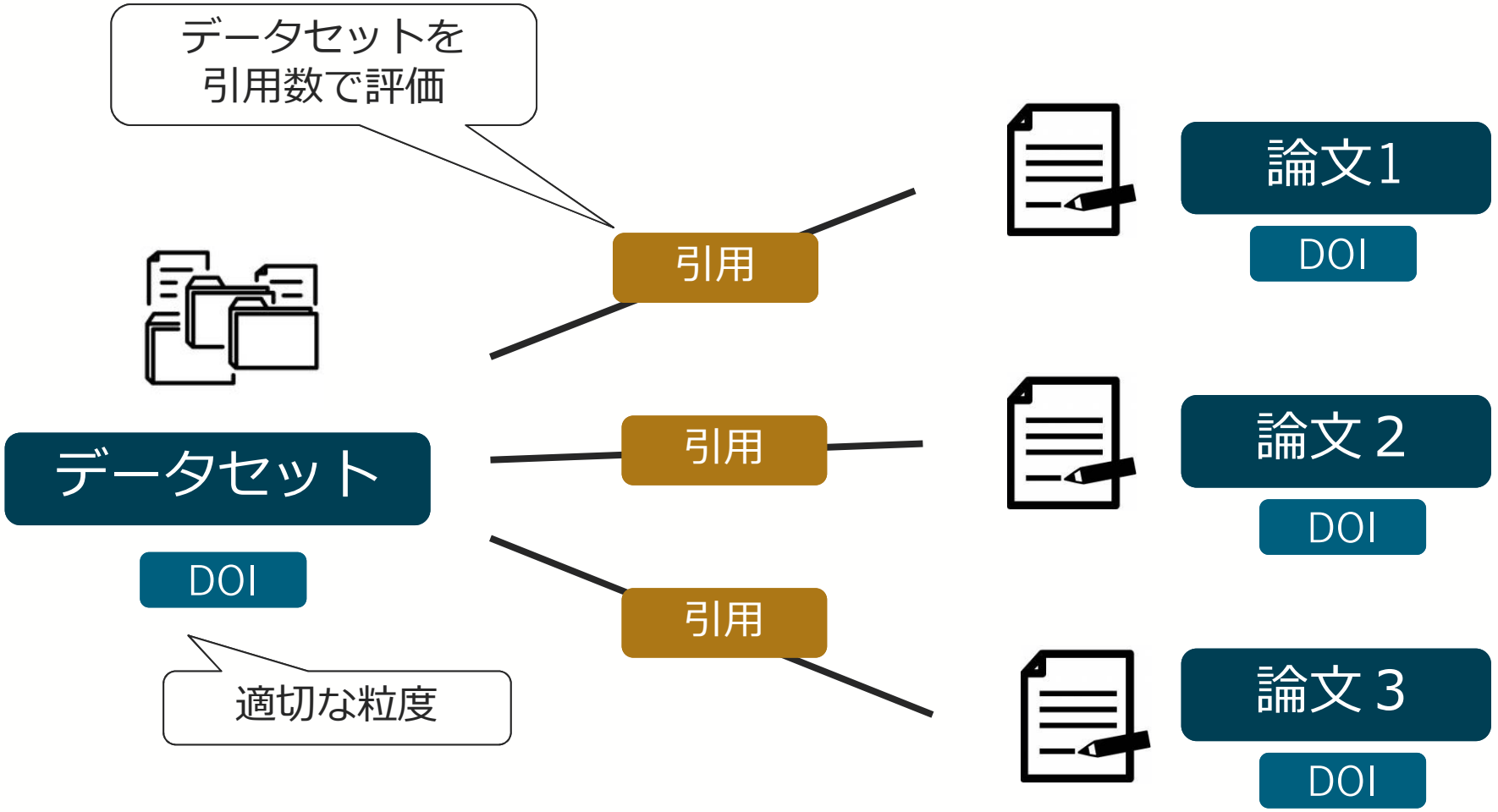
研究データの引用評価は、
論文と同じ方法で可能なのか？

引用された論文の評価の基本構造



論文は、論文単位でカウントをして評価出来るため、引用のカウントもシンプル

データに置き換えると…



・「粒度」とは、データ提供者が定義するまとまりのあるデータ単位のこと
・DOIは、「データ引用の単位」として「粒度」と置き替えられる

引用以外の粒度を決めるための考え方

National Institute for Environmental Studies, Japan

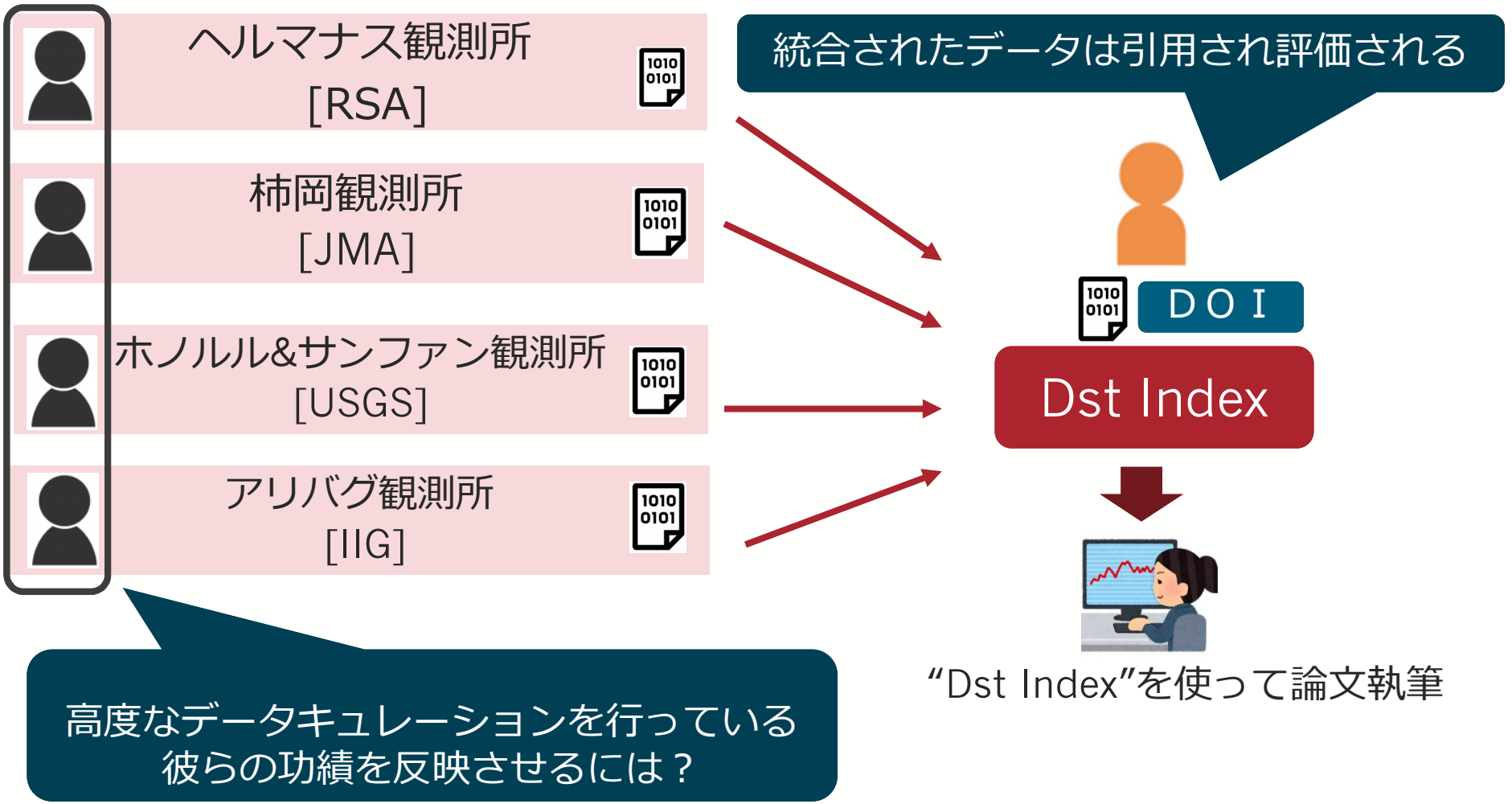
引用	データ提供者が引用を期待する粒度
データの性質	データセットとして意味のある粒度
アクセシビリティ	利用者がアクセスしやすい粒度
ID数	適切に運用できる数量にとどめた粒度

複数のデータを統合して得られるデータの場合

複数のデータを統合して得られるデータの問題

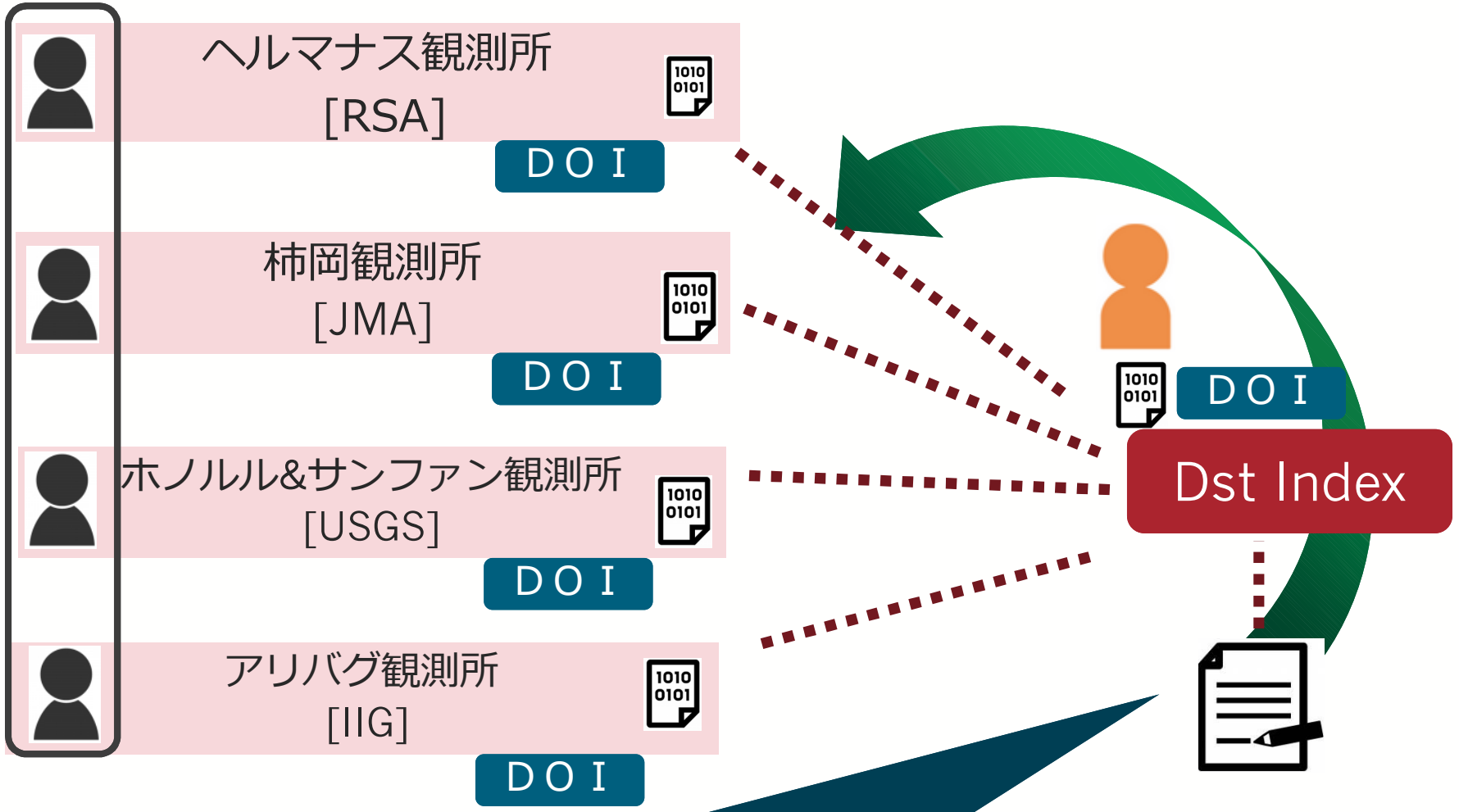
National Institute for Environmental Studies, Japan

Dst index (Disturbance storm time index)
複数の赤道付近で観測された平均的な地磁気変動量から算出した地磁気変動指数



村山他（2015）「観測データから生成された2次データに関するDOI引用法の新提案」を参考に図式化
http://www-kasm.nii.ac.jp/papers/takeda/15/murayama15sgepss.pdf

複数のデータを統合して得られるデータの問題



それぞれの観測所のデータにもDOIを付与し、
Dst Indexとリンクさせる。メタデータへの補記も必要。
遡って、評価できるようにする。

１次データと解析データが混在する場合

複次的なデータの生産工程

オリジナルデータとして価値が高い

1 次データ
（資源的データ）

- モニタリングデータ
- 社会調査の個票
- 調査記録/実験ノート

AIや機械学習で増加傾向

2 次データ
（解析工程のデータ）

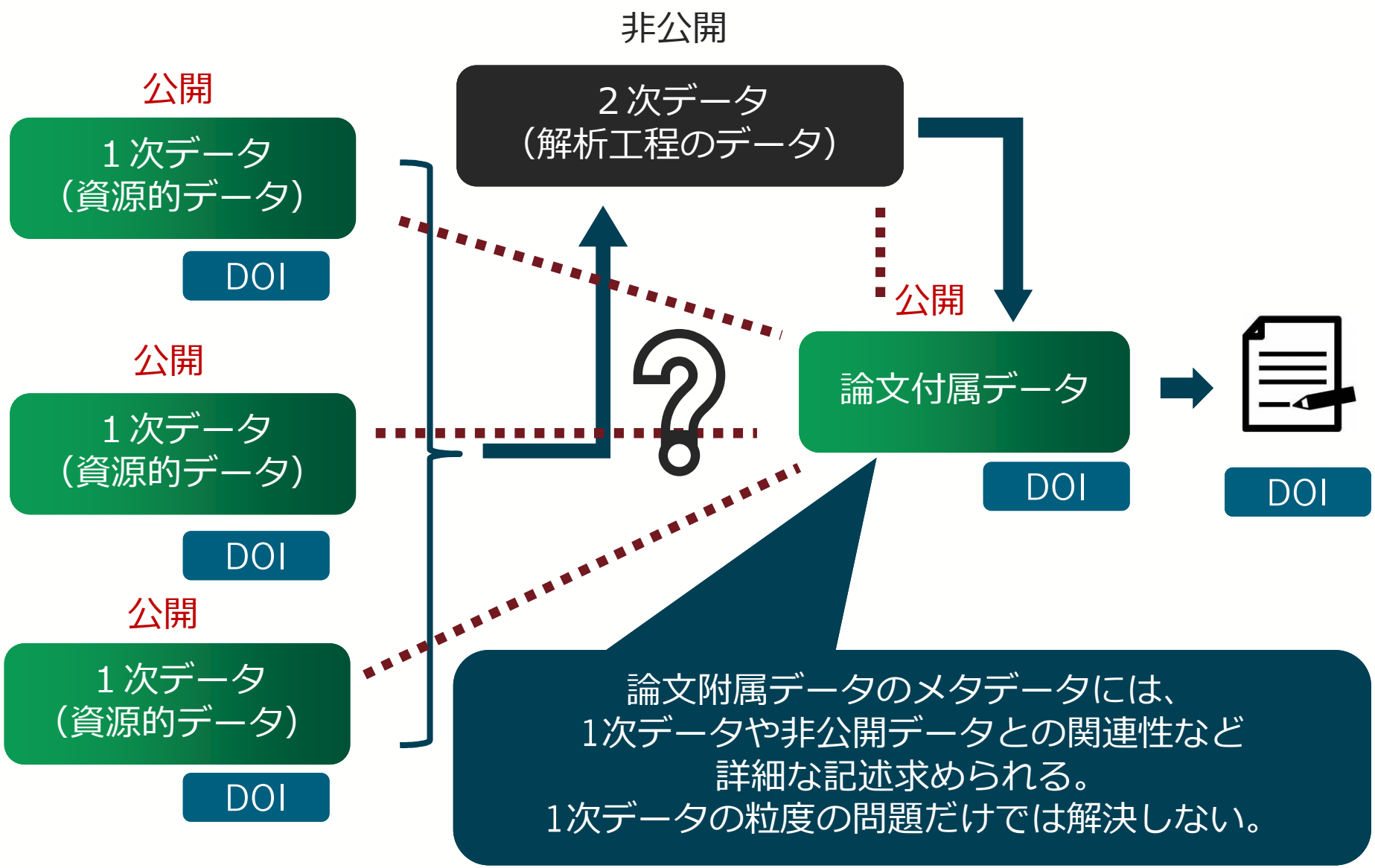
- 解析工程の結果データ
- 実験プロトコル
- プログラム/スクリプト

論文投稿に必須（になりつつある）

論文付属データ

- 論文で使用したデータセット
- 図表などSupplement

解析工程を挟むケース－附属データの附属情報



大規模な社会調査や疫学調査のデータの場合

比較的規模の大きい疫学調査の例

National Institute for Environmental Studies, Japan

子どもの健康と環境に関する全国調査

エコチル調査

出生コーホート調査

http://www.env.go.jp/chemi/ceh/results/index.html

調査機関：環境省　国立環境研究所
　　全国15地域の大学等に設置されたユニットセンター

大規模な共同調査

調査期間：2011年～2032年

22年もの年月

調査対象：約10万組の親子
　　（母親の妊娠中から子どもが13歳し達するまでの追跡的調査）

取得するデータ：診察記録、質問票、生体試料

データの種類が多い

主な研究利用：化学物質の影響を統計的に解析・分析

＊調査期間が長く、データ管理担当者の入れ替わりは必至
＊データの集計作業とその解析・分析が同時平行で実施

大規模な社会調査や疫学調査の問題

全項目をデータセットにして一つのDOIを付与するとした場合

項目１の集計　2010年 → 2011年に出版

項目２の集計　2011年 → 2012年に出版

後追いでのレファレンス修正は困難
↓
被引用のカウントから外れてしまう

項目３の集計　2012年 → 2013年に出版

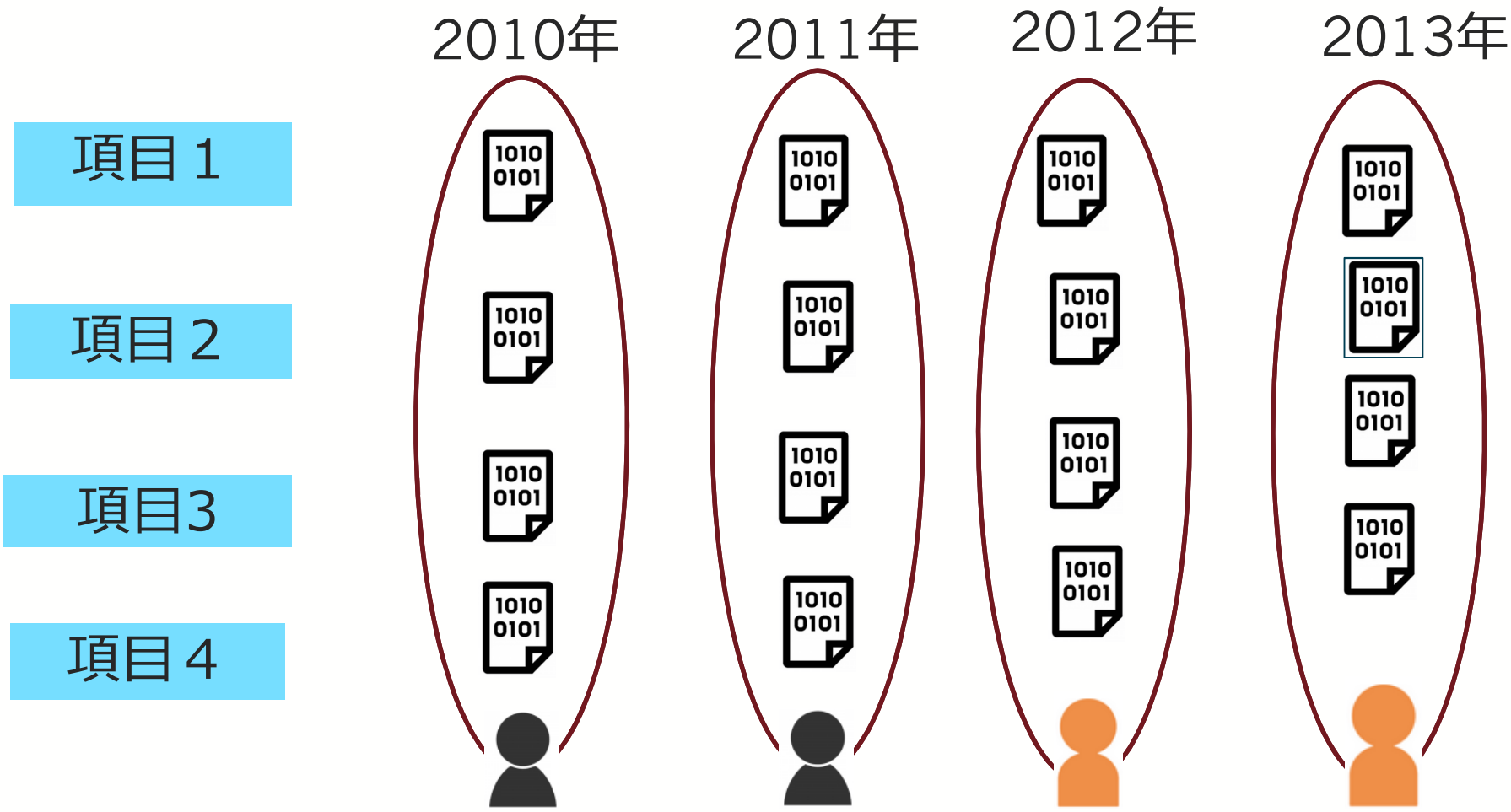
項目４の集計　2013年 → 2014年に出版

DOIを付与して引用OK

2013年に項目４の集計を終えて、ようやくDOI付与

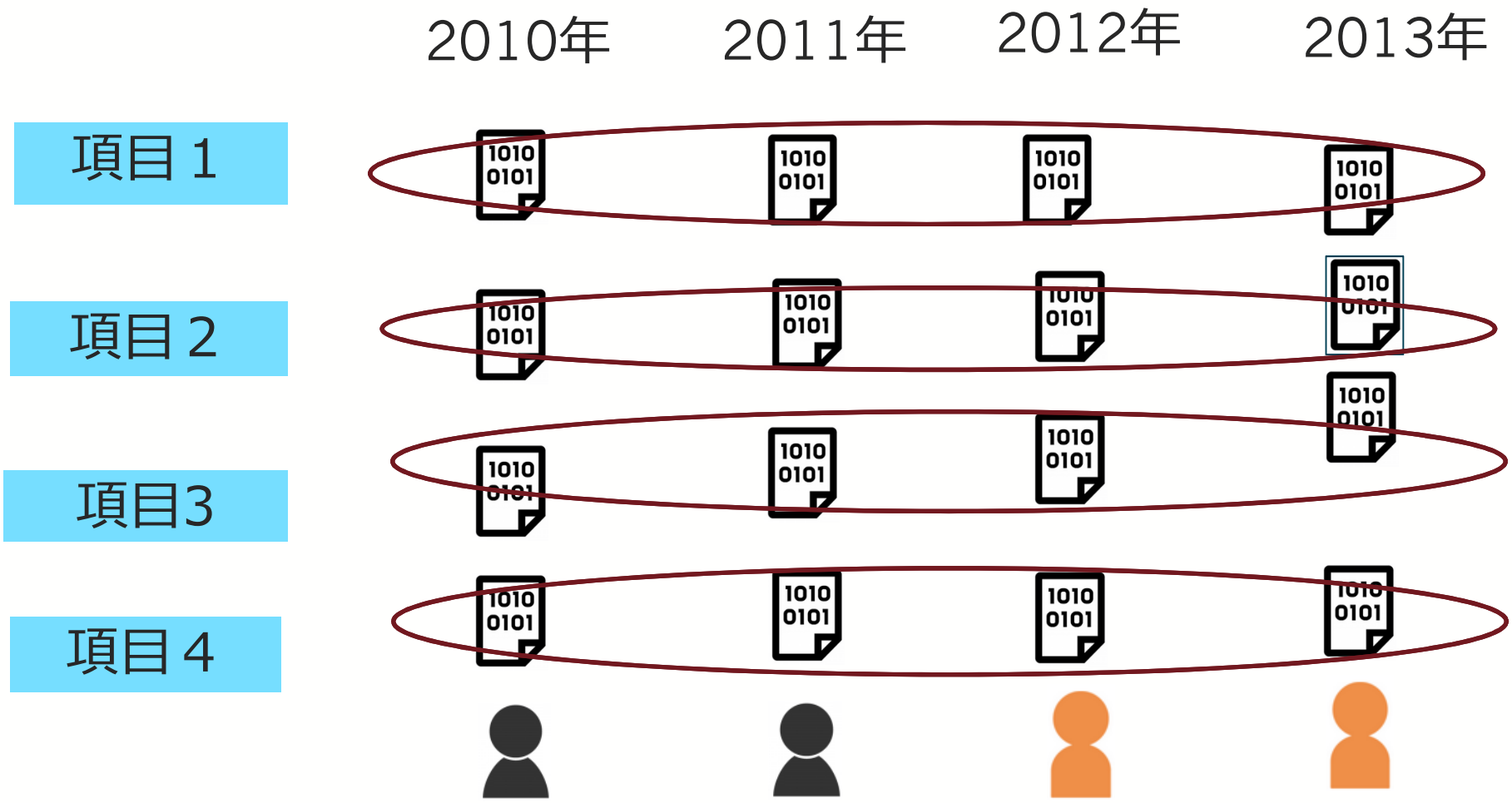
DOIの粒度は、公開時期も踏まえて考える必要がある

DOIの粒度を年次ごとにした場合



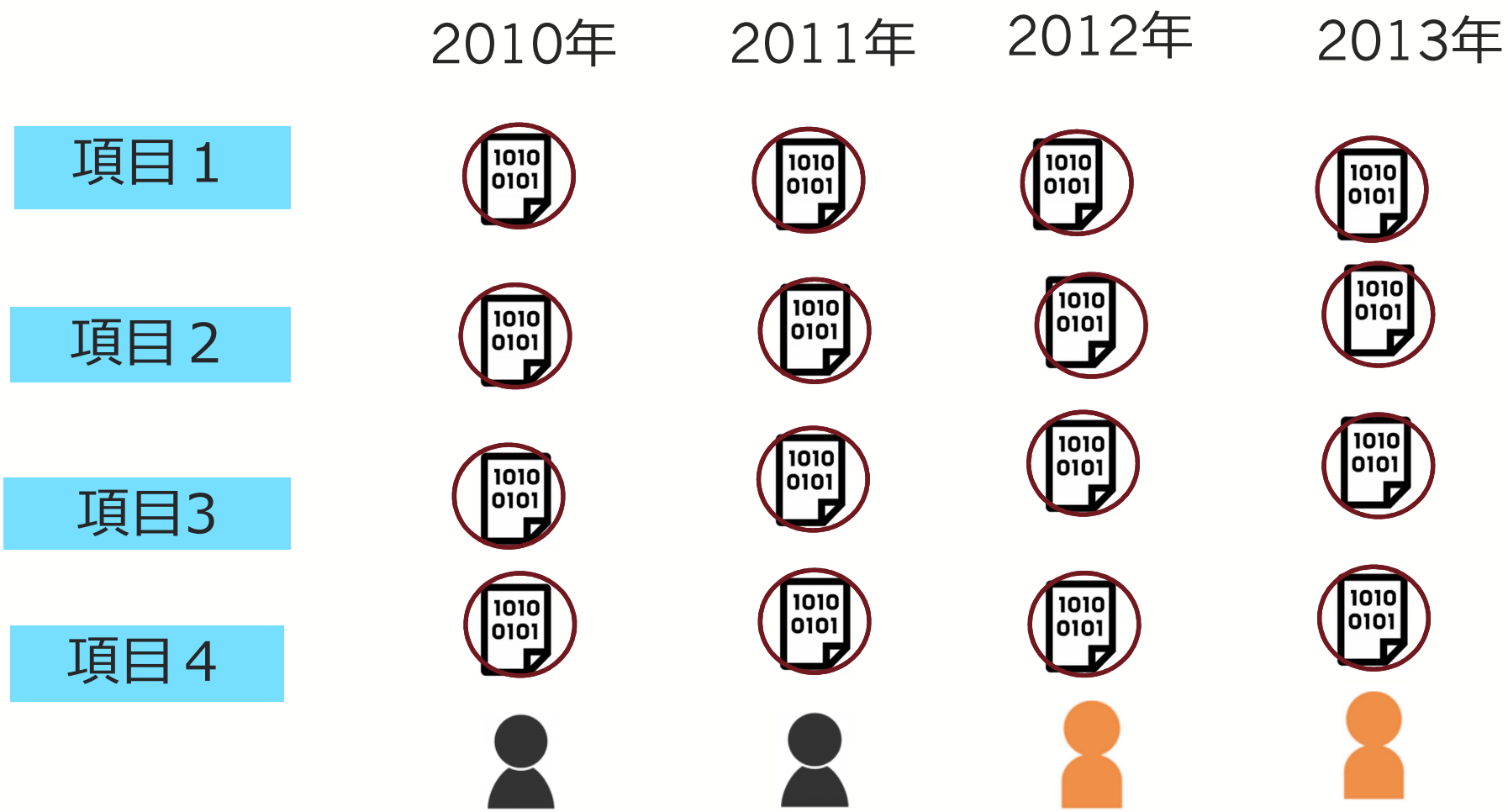
年次ごとにデータファイル形式やファイル名の規則を担当者間で統一させておくことが重要

DOIの粒度を項目ごとに場合



担当者が変わっても適切にバージョン管理出来るか

年次と項目両方で粒度を決めた場合



引用される際、適切なID数の粒度であるかどうか
利用者のアクセシビリティにも配慮する必要がある

まとめ

National Institute for Environmental Studies, Japan

- DOIの粒度の決定においては、論文等での引用を強く意識する必要があるのではないか。

- 研究データの引用・被引用による研究評価では、データ間関係をから評価出来るようなシステム設計が必要ではないか。

- 大規模な社会調査や疫学調査では、DOI付与のを粒度の観点から、研究データ管理計画策定当初の時点で、考慮しておく必要がある。

ご静聴ありがとうございました