

# 研究データの利用規約

## — 限界と解決案 —

真板英一(国環研)

2016-02-24

# 背景

- ・ 環境研・地球センター用に DOI Landing Page を作成中
- ・ 利用規約の内容を検討しているが、研究者との認識の違いが案外大きかった
- ・ その違いが生ずる背景について考察した

# 論点

- ・ 利用規約は学術エコシステムの一部である
  - ・ 最適な利用規約はシステム論的に決まる
  - ・ 利用規約だけを検討しても意味が無い
  - ・ 利用規約を補完する文化やITシステムも重要
- ・ 研究データの独特の性質
  - ・ 表現物ともソフトウェアとも違う
  - ・ CCライセンスは研究データに適するか？
- ・ 研究者の生態・習性・生存戦略を考慮すべき
  - ・ 特に、直接的コミュニケーションへの強い欲求
  - ・ 研究者は公務員と言うよりは起業家に近い（今日は省略）

# 研究者の意識調査

- Tenopir et al.(2015)
  - DataONE Usability WG の活動として調査実施
  - DOI: 10.1371/journal.pone.0134826
- Schmidt et al.(2016)
  - Belmont Forum の基礎調査として実施
  - DOI: 10.1371/journal.pone.0146695

# Tenopir et al.(2015)



RESEARCH ARTICLE

## Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide

Carol Tenopir<sup>1</sup>, Elizabeth D. Dalton<sup>2\*</sup>, Suzie Allard<sup>3</sup>, Mike Frame<sup>4</sup>, Ivanka Pjesivac<sup>5</sup>, Ben Birch<sup>6</sup>, Danielle Pollock<sup>7</sup>, Kristina Dorsett<sup>8</sup>

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0134826>

# 概要

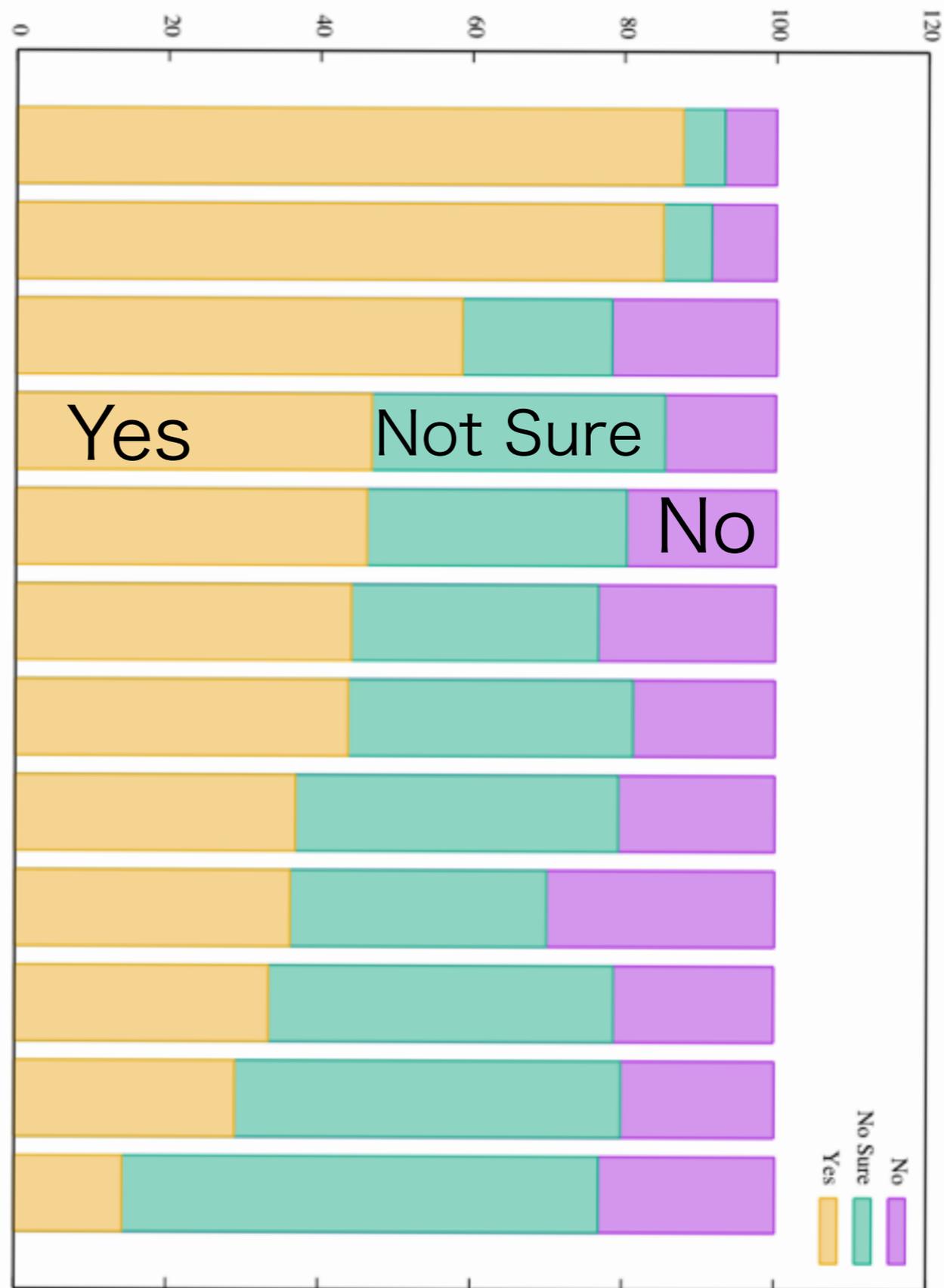
- ・ DataONE の Usability WG による調査
- ・ 雪だるま式アンケート（メール拡散方式）
- ・ 回答者数 1329人
- ・ 前報(Tenopir et al. 2011) との比較が目的

自分のデータの  
利用許可の条件として  
同意するもの

- ★ 共同研究の機会提供
- ★ 成果物の提出
- ★ 双方のデータの相互利用
- ★ 利用目的の事前説明
- ★ 成果物一覧の提出
- ★ 成果物の形態の事前確認
- ★ 共著
- 合法的な利用許可
- ★ 成果物の内容の事前確認
- データ作成費用の負担

★ 直接的なフィードバック、  
コミュニケーションの欲求

謝辞  
引用



Tenopir et al.(2015) S1 Appendix Table H を元に真板作成

# Schmidt et al. (2016)



RESEARCH ARTICLE

## Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey

**Birgit Schmidt<sup>1\*</sup>, Birgit Gemeinholzer<sup>2</sup>, Andrew Treloar<sup>3</sup>**

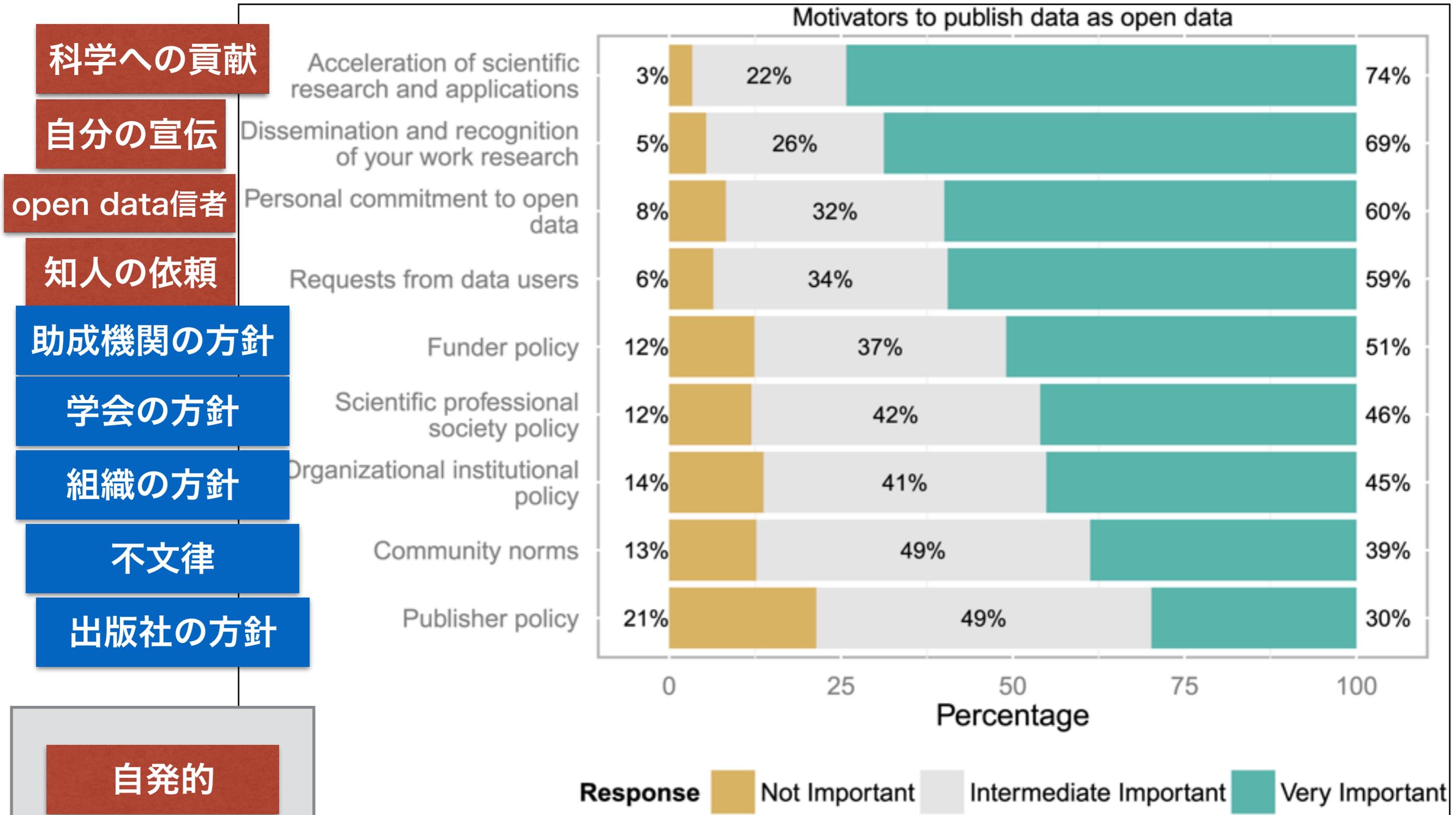
**1** State and University Library, University of Göttingen, Göttingen, Germany, **2** Systematic Botany, University of Giessen, Giessen, Germany, **3** Australian National Data Service, Melbourne, Australia

<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146695>

# 概要

- ・ オープンサイエンスに対する、研究者の意識調査
- ・ Belmont Forum にインプットするための調査
- ・ Web アンケート形式
- ・ 回答者数 1330人
- ・ オープンサイエンスに理解がある人々に偏っている

# データ公開の動機



科学への貢献

自分の宣伝

open data信者

知人の依頼

助成機関の方針

学会の方針

組織の方針

不文律

出版社の方針

自発的

外圧

Fig 6. Motivators to publish data as open data.

# データ公開の障壁

先に論文化

法規制

原典の忘却

誤解誤用の不安

知的財産保護

組織の制約

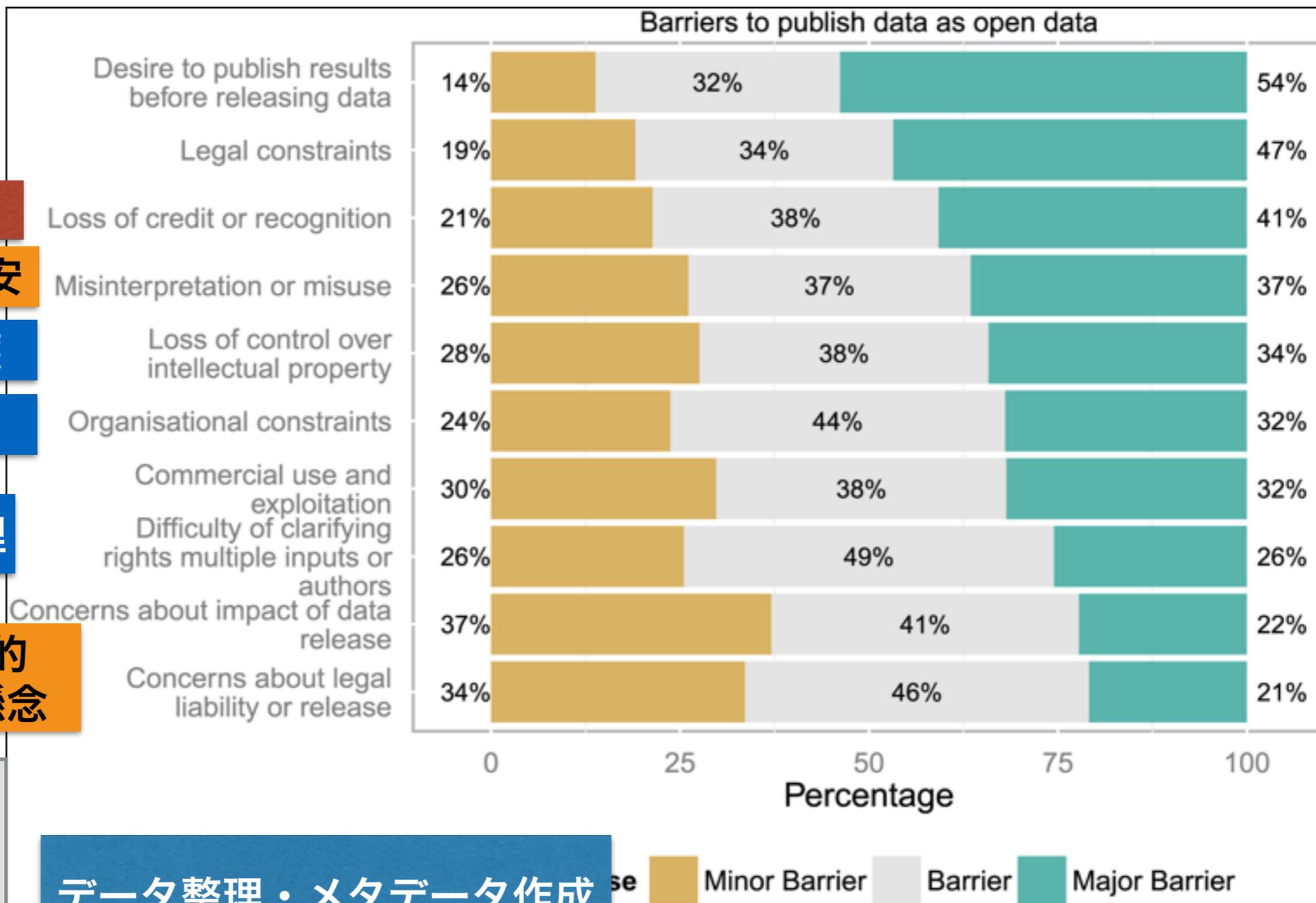
権利関係の整理

社会的・法的  
トラブルの懸念

個人的利害

不安

制度的理由



データ整理・メタデータ作成  
のコストへの言及がない  
(調査票に選択肢がなかった)

Fig 7. Barriers across countries.

# 研究者の意識調査

- ・ 研究者の個人的利害

- ・ 先に論文を出したい

単なるエゴイズムに過ぎないのか？

- ・ 直接的なフィードバックが欲しい

今日の主要論点

- ・ 心理的な不安

- ・ 誤解・誤用されるのではないか

- ・ 何かトラブルに巻き込まれるのではないか？

杞憂に過ぎないのか？

# フィードバック

- ・ 研究者がフィードバックを欲しがるのは何故か？
  - ・ 「研究」は常に発展させなければならないから
- ・ データの再利用自体は自分の研究に影響しない
  - ・ 引用されて業績は増えるかも知れないが、自分自身の研究の役には立たない

学術的価値

研究者は縦展開が好き

引用

更に発展

フィードバック

他の研究者

発展

?

データ公開

二次的価値

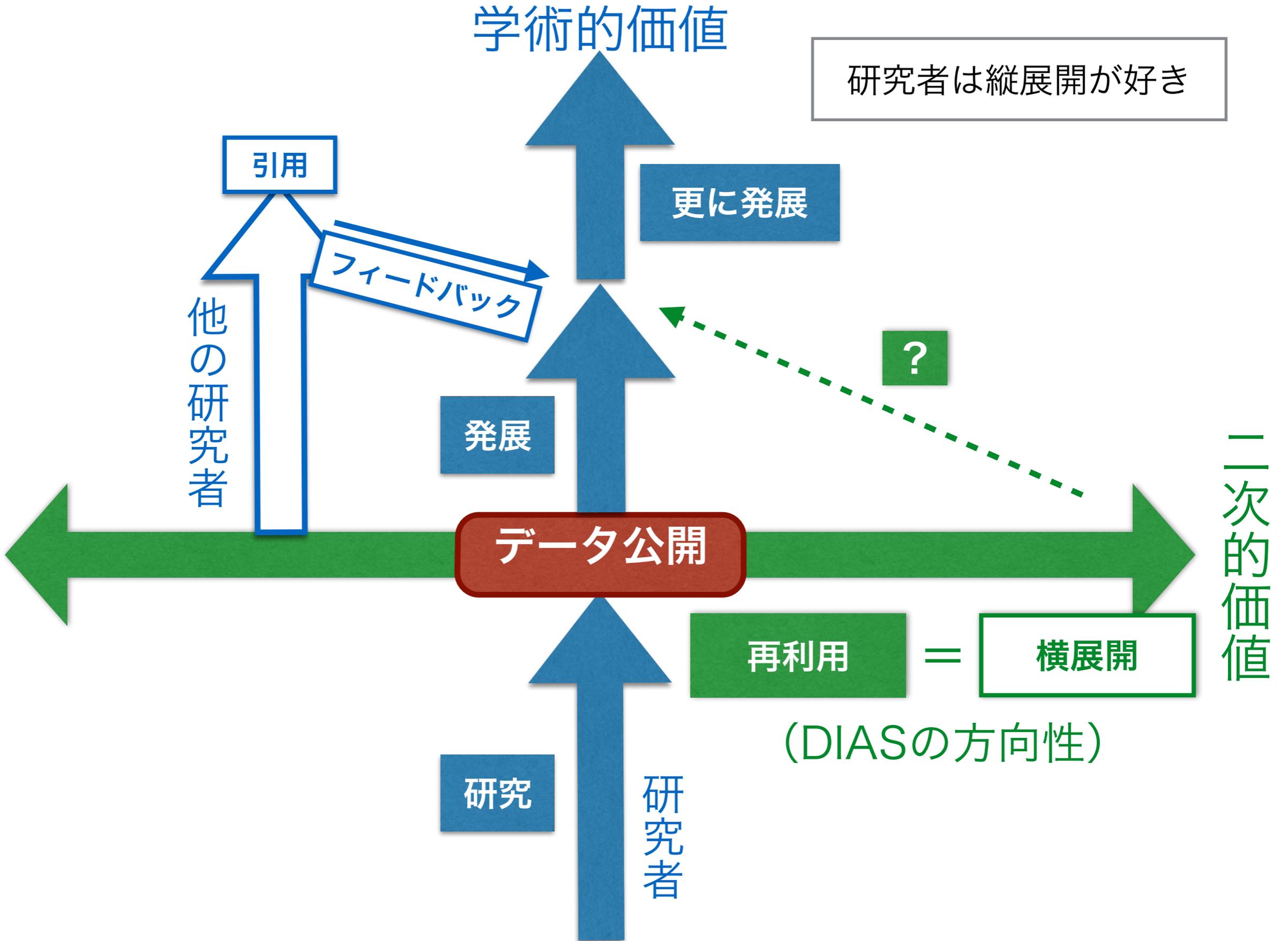
再利用

横展開

(DIASの方向性)

研究

研究者



# データ引用の有効性

- ・ Data citation はデータ公開のインセンティブになるのか？
- ・ 被引用回数の増加は、目的か？ 結果か？
  - ・ 業績を挙げることが目的で、研究はその手段だというなら、data citation はインセンティブになる。
  - ・ 楽しく研究することが目的で、業績は単なる結果だというなら、data citation は直接的なインセンティブにはならない。
- ・ オープンサイエンスが想定する研究者像は、どちらのタイプなのか？

# データ引用の戦略

利用規約を緩くして、  
「再利用＝被引用回数」  
を増やす戦略

昨今のオープンサイエンスの  
議論ではこちらが主流か。  
オープンサイエンス推進派の  
基本的立場。

「再利用＝被引用回数」  
を犠牲にしても、  
直接的フィードバックを  
確保する戦略

現場の研究者の心情（習性？）に  
近い考え方か。  
研究者が利用規約をきつく  
したがることの背景。

# フィードバック

- ・ 研究者のエゴイズムだ、とは言い切れない
  - ・ 学術研究の「常時発展」的性質を考えると
- ・ データ提供者への直接的フィードバックを担保する必要がある
  - ・ ただ、それを利用規約で実現するのか、他の手段を用いるか、は議論の余地がある
  - ・ 個人的には、他の手段で実現する方が良いと思う

# CC と OSS



## FAQ よくある質問と回答

- ソフトウェアにクリエイティブ・コモンズ・ライセンスを付与することができますか？

可能ですが、お勧めはできません。

<http://creativecommons.jp/faq/#a6>

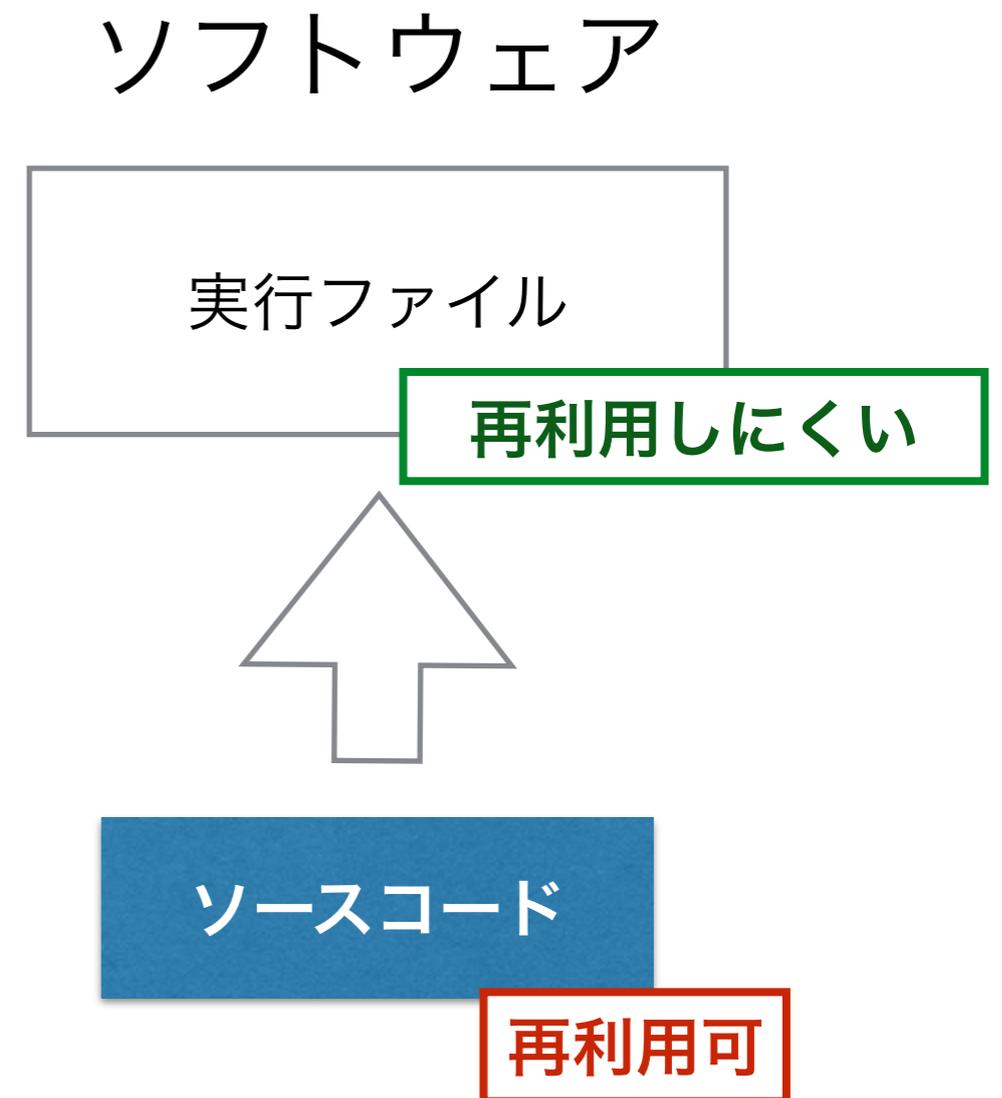
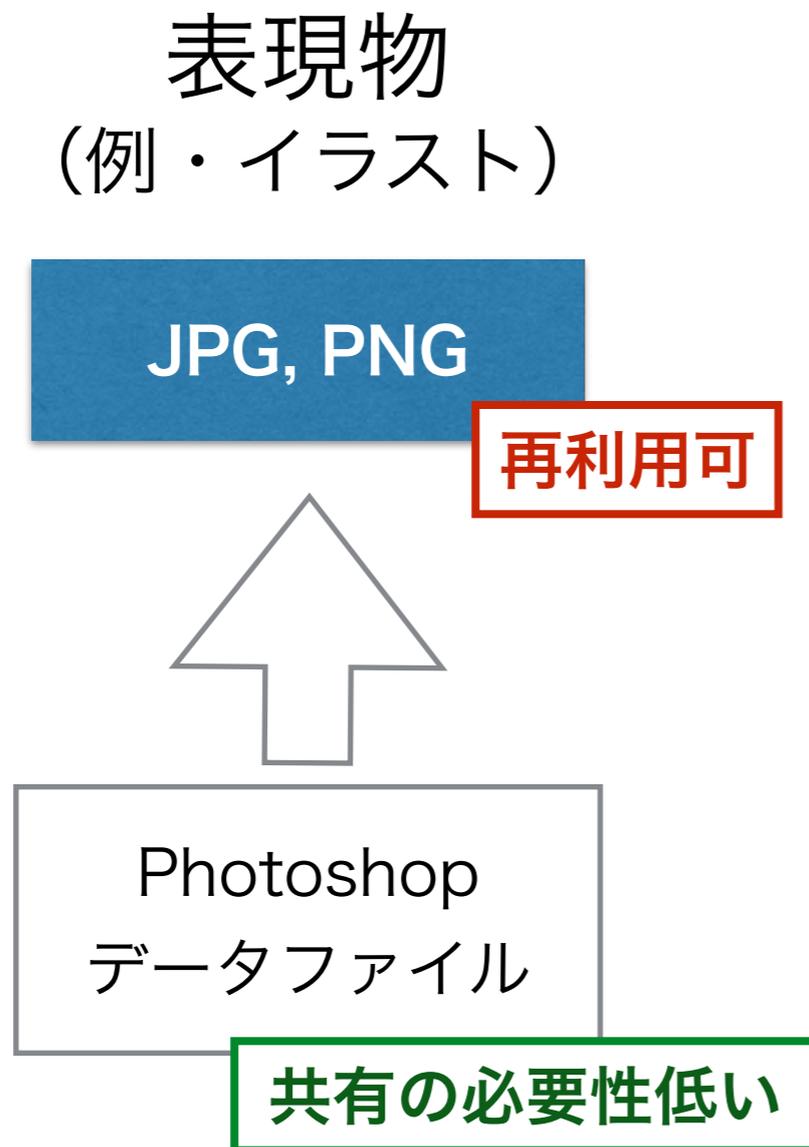
# CC と OSS

## CCライセンスがソフトウェアに適さない理由

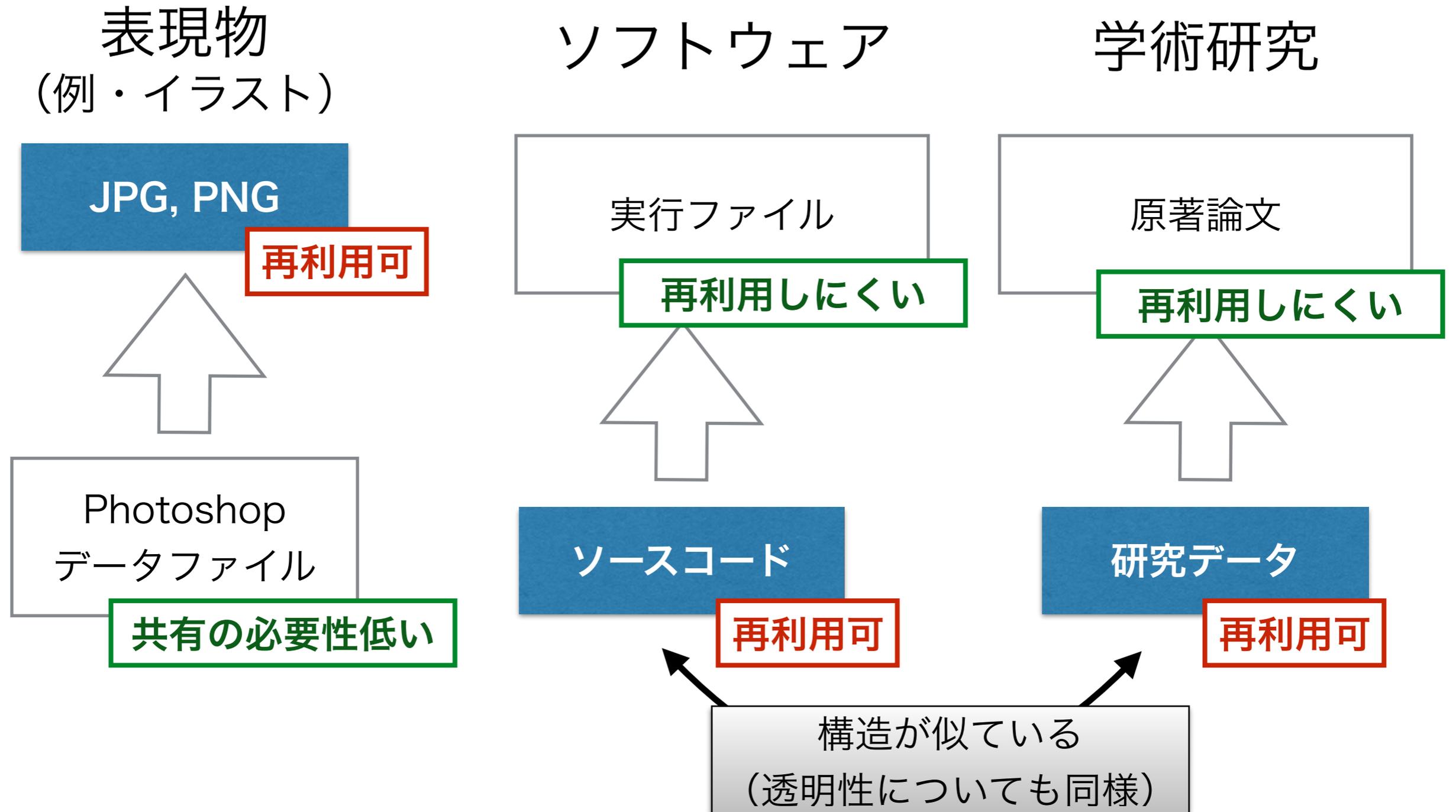
- ・ ソースコードという概念がない
- ・ パテントに関する規定がない
- ・ 既存のソフトウェア用ライセンスとの互換性がない

<https://creativecommons.org/faq/#can-i-apply-a-creative-commons-license-to-software>

# 再利用性の違い



# 再利用性の違い



# 多様性 vs 正しさ

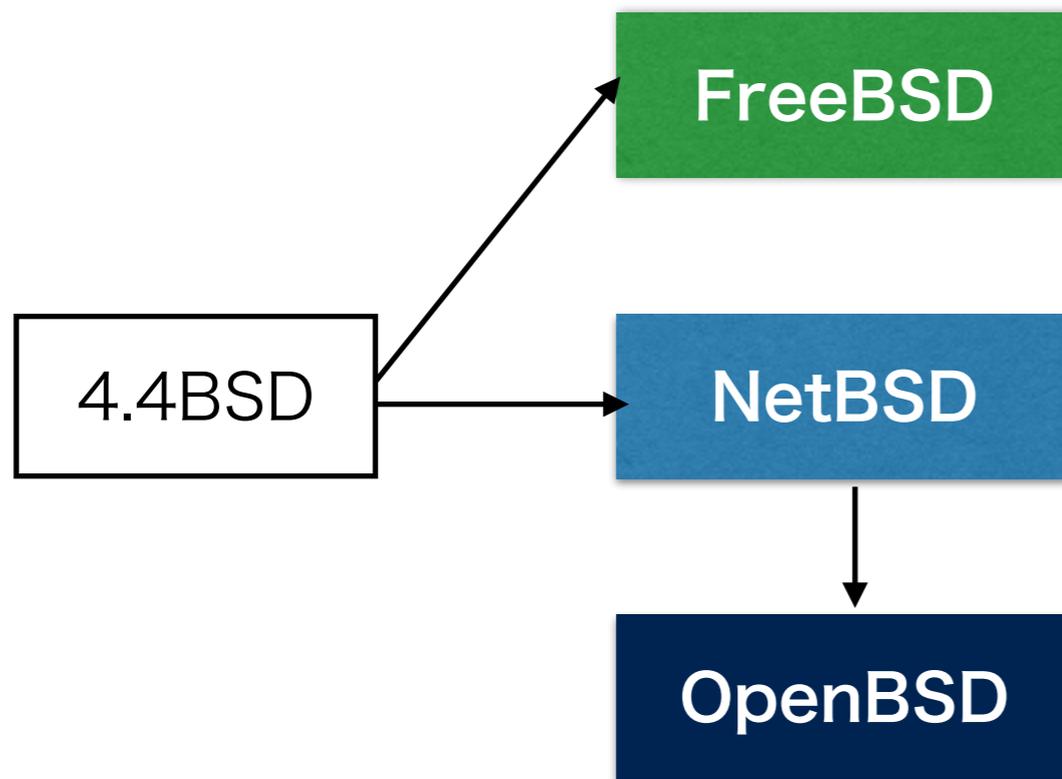
- ・ 表現には基本的に「正しい・間違い」は無い
  - ・ 多様性に価値がある
- ・ ソフトウェアには「正しい」「良い」がある
  - ・ バグ（仕様に対する実装の間違い）
  - ・ 機能の多さ、処理効率の良し悪し
  - ・ ただし、多様性も尊重される

# ソフトウェアの多様性

BSD系OSの例

(価値観)

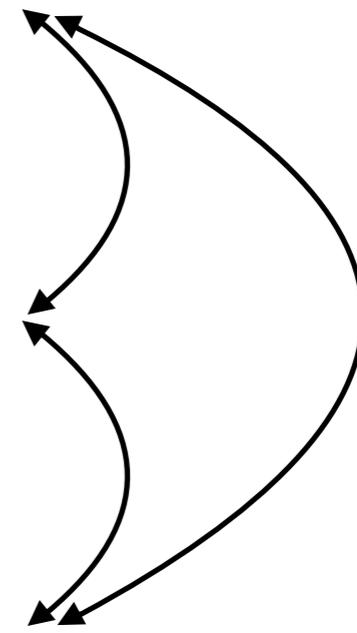
良いところは  
どんどん真似る



利便性

移植性

安全性



Open & Free ライセンス

開発コミュニティの価値観の多様性



ソフトウェアの多様性

# ソフトウェアの正しさ

Bug Report/Bug Track

FreeBSD Bugzilla - Bug List

Home | New | Browse | Search |  Search [?]

[Hide Search Description](#)

**Status:** New, Open, In Progress, UNCONFIRMED

This result was limited to 500 bugs. [See all search results for this query.](#)

ID	Product	Component	Assignee	Status
<a href="#">2090</a>	Base System	bin	freebsd-bugs	In Prog
<a href="#">2298</a>	Base System	kern	freebsd-bugs	In Prog
<a href="#">4116</a>	Base System	bin	freebsd-bugs	In Prog
<a href="#">5031</a>	Base System	bin	freebsd-bugs	In Prog

コミュニティからの  
フィードバックで担保

協力して良いものを作ろう  
という文化の共有

Pull Request

GitHub

maitaA1 / firsttest

Code Issues 0 Pull requests 1 Wiki Pulse Graphs

Dev #5

Open maitaA1 wants to merge 2 commits into master from dev

Conversation 0 Commits 2 Files changed 1

maitaA1 commented just now

pull request test

maitaA1 added some commits 3 minutes ago

for pull request

Merge branch 'master' of https://github.com/maitaA1/firsttest into dev

Add more commits by pushing to the dev branch on maitaA1/firsttest.

This branch has no conflicts with the base branch  
Merging can be performed automatically.

Merge pull request You can also open this in GitHub Desktop or view command line

Write Preview AA B i « > ☰

# ライセンス + 文化

OSSではライセンスと文化は相補的関係

ライセンス

制限を緩くして、  
自由な再利用を促進

多様性

文化

コミュニティを作り、  
フィードバック文化を育成

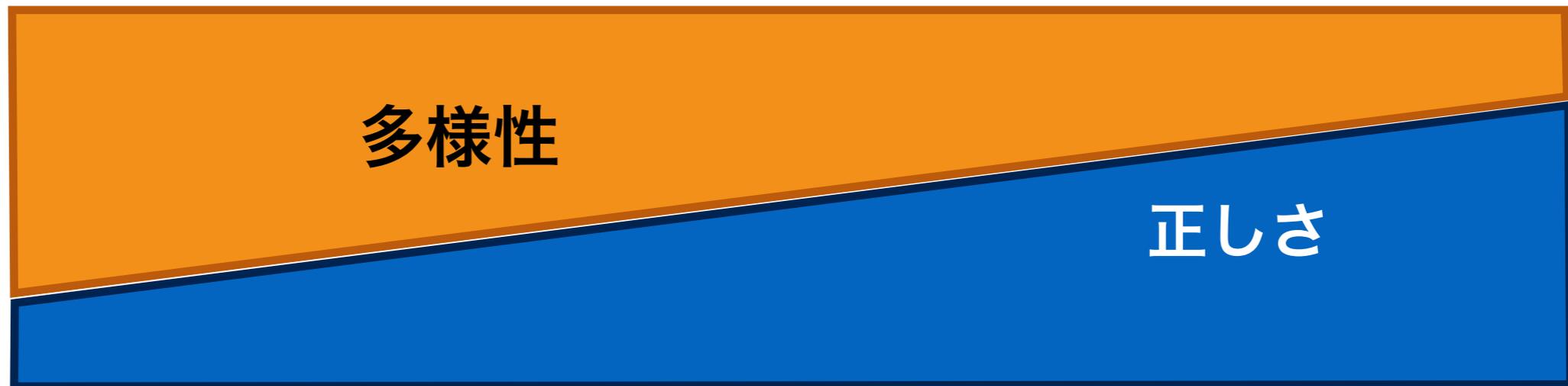
正しさ

# 研究データの特徴

表現物

ソフトウェア

研究データ



研究データは、精度や正確性が最も重要。

多様性は、むしろ障壁である場合が多い？  
実験条件が違ったりすると、  
相互比較や統合解析・メタ解析が難しくなる。

# 研究データの正しさ

## 研究データの正確性をどのように担保するか

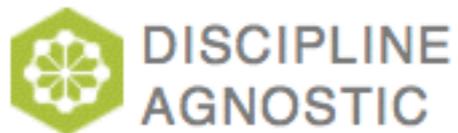
- ・ 従来: データ作成者が心血を注いで QC する
- ・ オープンサイエンス時代に相応しい方法がある？

# Social Curation



<https://dataconservancy.org/>

Lib, Johns Hopkins Univ.  
US DataNet Prog. のひとつ



Users of a DC instance will have access to a cross-disciplinary data curation system that facilitates data discovery and integration. Data within an instance can be broadly advertised using a variety of discipline specific and industry standard mechanisms, making it visible in a variety of external catalogs and registries, greatly expanding the user base for data stored within the instance.

同じデータに、  
様々な分野の研究者が  
リモートで  
キュレーションできる  
リポジトリシステム  
らしい。  
(詳細は未調査)

<https://dataconservancy.org/software/>

# 研究者SNS

例

ResearchGate

<https://www.researchgate.net>

- ・ 論文（のプレプリント）のリポジトリ
- ・ SNS機能（論文のリクエストなど）
- ・ 求人マッチングサービス

これでマネタイズ  
しているベンチャー

（日本で言うと、researchmap + JREC-IN Portal か？）

今のところ、研究データ用の機能は無いが、  
将来的に、SNS的なシステムが公開データの  
フィードバックにも応用できるかも

# リポジトリの機能強化

研究データリポジトリに OSS的な

- ・ Issue Tracking System
- ・ Pull Request 機能

などがあれば、利用者がフィードバックしやすい？

# 研究データのコミュニティ

- ・ 研究者はフィードバックを欲しがる。
- ・ 今はフィードバックの文化やITシステムが無いので、利用規約に色々書きたがる。
- ・ 文化・システムがあれば、「利用規約はゆるゆるでいいや」と考えるようになる（と思う）。
- ・ フィードバックが発生するようなコミュニティを育成する必要があるのではないか。

# まとめ

- ・ データ提供者の欲求はライセンスだけでは解消しない
  - ・ ライセンス + フィードバック文化(+ICT)
- ・ 研究データ専用のライセンスが必要？
  - ・ CCライセンスでは不十分のような気がする
- ・ オープンサイエンス時代のコミュニケーションとは？
  - ・ 誤解誤用の防止にも役立つかも知れない