

「巨人の肩」と データサイテーション

国立情報学研究所 / 総合研究大学院大学
北本朝展（きたもとあさのぶ）

<http://agora.ex.nii.ac.jp/~kitamoto/>

自己紹介



@KitamotoAsanobu

- 情報学の最前線（他分野との境界領域）において、**データを基盤に価値を生み出す研究活動**を行っている。
- 特に、**地理的データ、視覚的データ**を中心に、**検索・解析基盤**を構築する。
- **地球科学・人文科学**を中心に、**他分野の研究者と積極的に協働**を進めている。

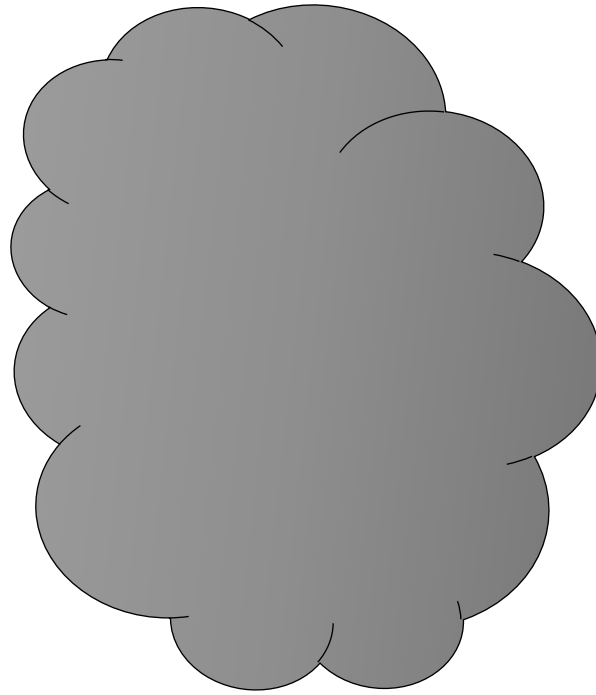
オープンサイエンスへの のコンバージェンス

私の定義

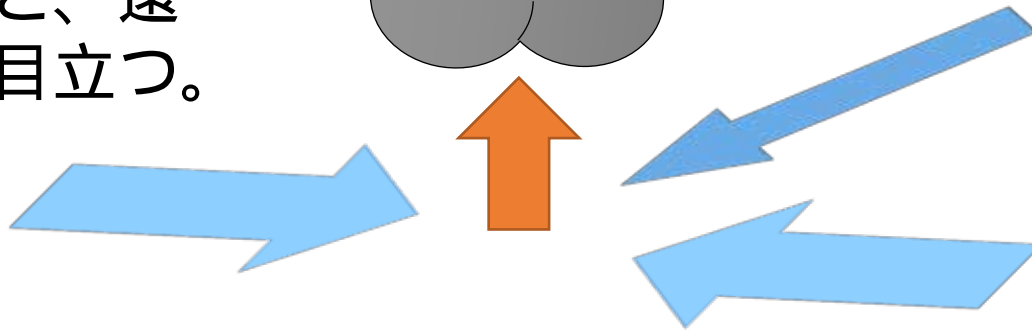
オープンサイエンスとは、従来の学術研究に対する不満を解決するために、オープンという概念に希望を見出した人々が集う、「夢の収束点」である。

オープンサイエンスへの収束

様々な動きが収束する場所に、「オープンサイエンス」という積乱雲がモクモクと立ち上がってきた。雲が空高く成長すればするほど、遠くからでも目立つ。



個々の人々はみんな違う夢を見ている（同床異夢）。でも、全体の流れを見ると、希望はなんとなく一か所に集まっている。



オープンデータとは？

<https://okfn.org/opendata/>

Key Features of Openness

- Availability and access
- Reuse and redistribution
- Universal participation

Why Open Data?

- Transparency
- Releasing social and commercial value
- Participation and engagement

「**利便性**」と「**透明性**」と「**参加**」の3軸のどれを重視するか。オープンの狙いによって異なる。

オープンサイエンス因数分解

	利便性	透明性	参加
オープンデータ			
オープンアクセス			
オープンイノベーション			
オープンガバメント			
シチズンサイエンス			
研究再現性			
データサイテーション			

詳しくは以下のスライドで

Convergence to Open Science

オープンサイエンスへのコンバージェンス

Formation of a Community to Foster
Shared Perception from Different Dreams
同床異夢から共通認識を醸成するコミュニティの形成

National Institute of Informatics / SOKENDAI

Asanobu KITAMOTO

<http://agora.ex.nii.ac.jp/~kitamoto/>

2015/09/17

オープンサイエンスデータWS

1

オープンサイエンス
データ推進ワーク
ショップ
2015年9月17日～18日
@京都大学

<http://agora.ex.nii.ac.jp/~kitamoto/research/publications/osd15.html.ja>

「巨人の肩」とデータ サイテーション

データサイテーション

不満

- データの整備・運用は、手間がかかる研究の割にはきちんと評価されない。
- データ研究の評価のために新しい枠組みを作るのは、大変な仕事だし、研究者も付いてこない。

夢

- データ論文とデータ引用の組み合わせにより、既存の枠組みに乗っかる形でデータ研究を評価する。
- 「引用」は研究者にも馴染みの方法であり、導入への心理的障壁は低い。

データジャーナル

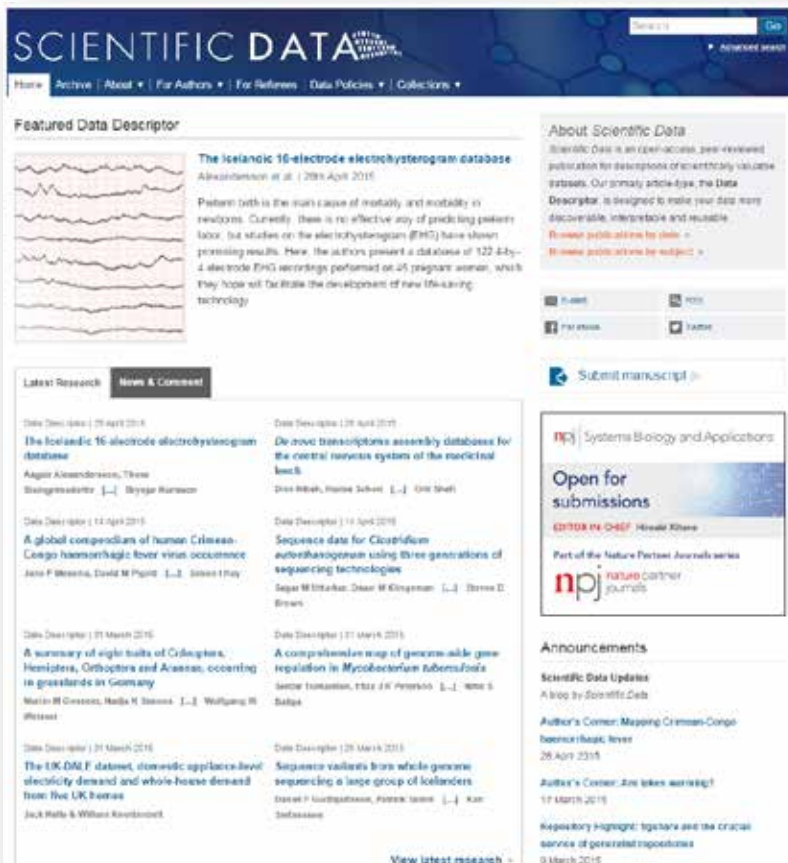
不満

- データは論文になりにくいいため、データの詳細に関する正式な文書がない。
- データはクオリティもバラバラ。再利用したくても、どれが良いデータなのかわからない。

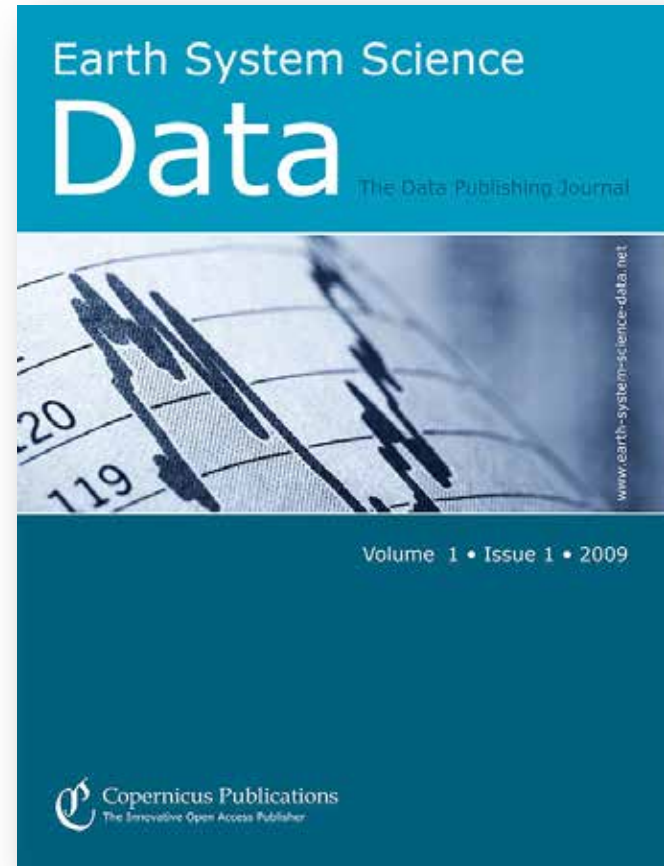
夢

- データに関する記述をデータ論文として認め、これを正式なメタデータとする。
- データ論文として査読を行い、きちんと作られたデータを区別可能とする。

データジャーナルの例



Scientific Data (Nature publishing group)

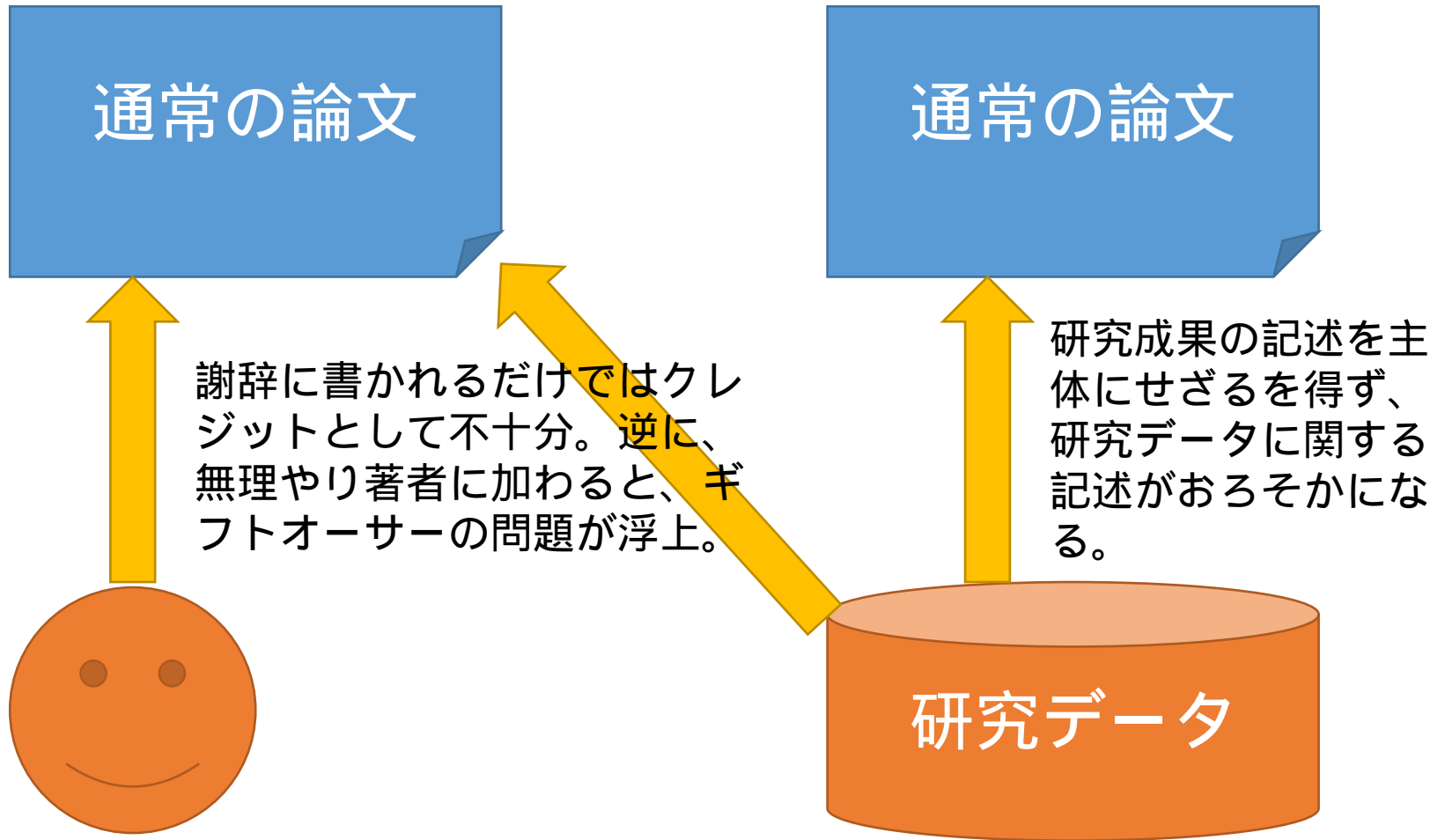


Earth System Science Data (Copernicus)

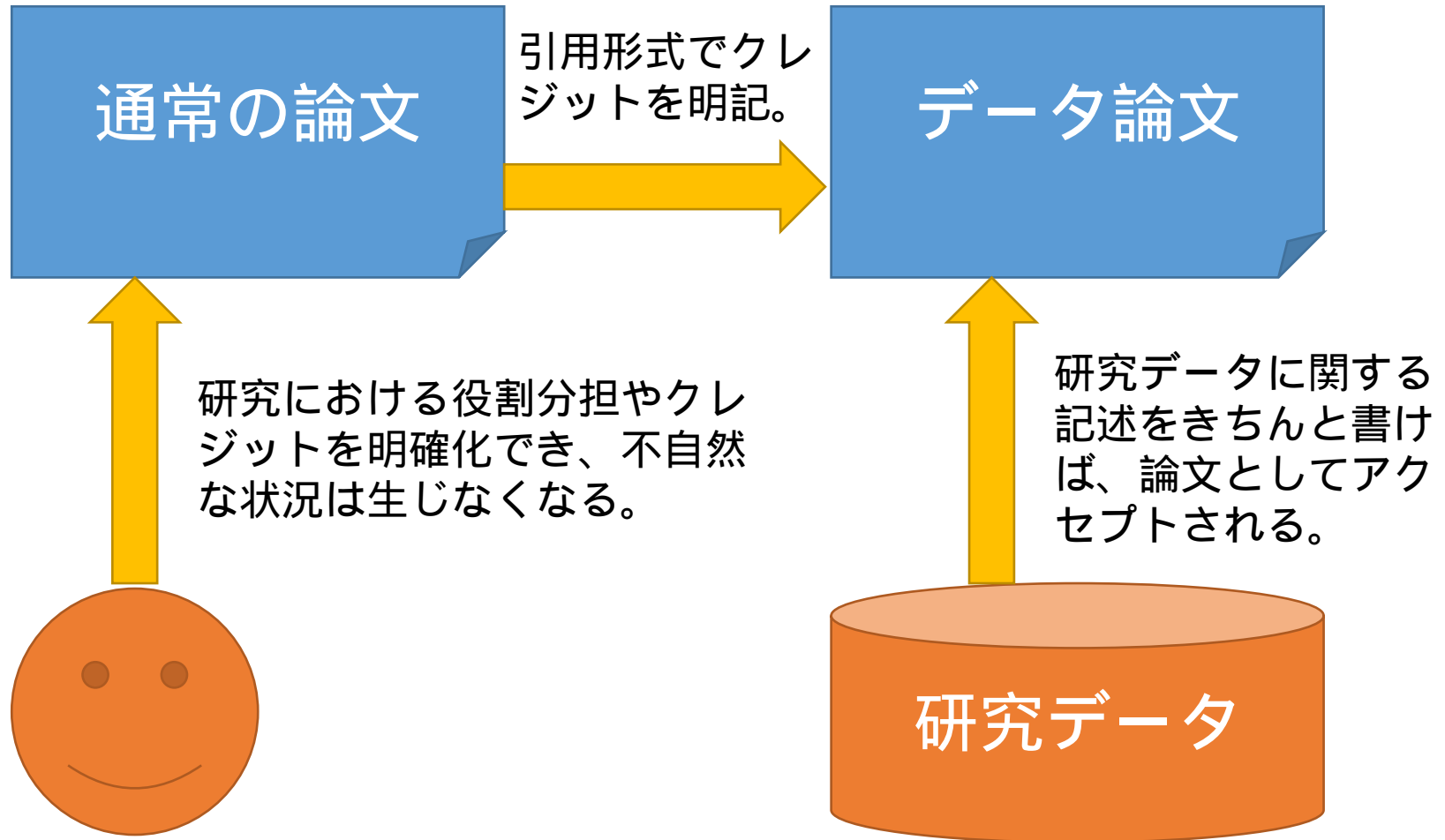
データジャーナルの査読

- Scientific Dataの例では基準は以下の通り。
 1. Experimental Rigour and Technical Data Quality
 2. Completeness of the Description
 3. Integrity of the Data Files and Repository Record
- データから生まれる発見の重要性や新規性は問わない。著者によるデータの解釈も不要。
- データ論文は、データの生成方法、品質管理、データ生成プロセスの厳密さなどを査読し、きちんとしたデータの再利用を促進するもの。

従来の学術出版



データ論文を含む学術出版



研究の価値と「巨人の肩」



巨人の肩の上に立つ

Stand on the shoulders of giants

- 学術研究は、先人の成果の上に立つことで、より先を見通せるようになる。
- 「巨人の肩」を作った人は偉い人である、はずだが、現在そうなっているか？

研究の網羅的な評価へ

- **なんでも引用**：すべての研究成果が、**引用という形式**でクレジットを受ける **既存の引用プラットフォーム**を使った、研究成果の網羅的な評価が実現する。
- **なんでもID**：すべての研究貢献が、**永続的なIDを結ぶ網の目**（有向グラフ）として表現される 従来型の論文評価に限定されない**多様なメトリックス**を使った、研究貢献の網羅的な評価が実現する。

研究成果へのIDの付与

- **DOI** : 10.1241/johokanri.56.881
 - **DOI-Prefix** : 10.1241 一意性はIDFが保証
 - **DOI-Suffix** : johokanri.56.881 一意性は各レポジトリが保証しなければならない
 - **Locator** : http://doi.org/10.1241/johokanri.56.881
- **Persistent identifier** : 永続的に一意性を保証するような識別子の体系。
 - **Actionable Locator** : ウェブ上で識別子をレゾルブし、（最新の）所在を伝える仕組み。

ソフトウェアへのIDの付与

The screenshot shows the GitHub Guides page for 'Making Your Code Citable'. The page has a dark blue header with the title and a '10 minute read' indicator. The main content area is white with a blue sidebar on the right. The sidebar contains a table of contents with links to 'Intro', 'Choosing Your Repo', 'Login to Zenodo', 'Check Repo Settings', 'Create a New Release', 'Minting a DOI', and 'Finishing up'. The main text explains that Digital Object Identifiers (DOIs) are the backbone of the academic reference and metrics system and provides instructions on how to archive code on GitHub and assign a DOI. A 'Pro Tip' section offers advice for researchers. The 'Choose your repository' section explains that repositories are the basic element of GitHub and provides instructions on how to choose one. The 'Login to Zenodo' section is partially visible at the bottom.

<https://guides.github.com/activities/citable-code/>

The screenshot shows the Zenodo website for a publication titled 'Analysis of Cartesian Genetic Programming's Evolutionary Mechanisms' by Brian W. Goldman. The page features a blue header with the Zenodo logo and the tagline 'Research. Shared.'. The main content area is white with a blue sidebar on the right. The sidebar contains a table of contents with links to 'Intro', 'Choosing Your Repo', 'Login to Zenodo', 'Check Repo Settings', 'Create a New Release', 'Minting a DOI', and 'Finishing up'. The main text explains that Digital Object Identifiers (DOIs) are the backbone of the academic reference and metrics system and provides instructions on how to archive code on GitHub and assign a DOI. A 'Pro Tip' section offers advice for researchers. The 'Choose your repository' section explains that repositories are the basic element of GitHub and provides instructions on how to choose one. The 'Login to Zenodo' section is partially visible at the bottom.

<https://zenodo.org/>

研究を取り巻く環境のID化

研究者や研究助成機関と紐づけ

ORCID
Connecting Research and Researchers

FOR RESEARCHERS | FOR ORGANIZATIONS | ABOUT | HELP | SIGN IN

DISTINGUISH YOURSELF IN THREE EASY STEPS

ORCID provides a persistent digital identifier that distinguishes you from every other researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between you and your professional activities ensuring that your work is recognized. [Find out more.](#)

- 1 REGISTER** Get your unique ORCID identifier. Register now! Registration takes 30 seconds.
- 2 ADD YOUR INFO** Enhance your ORCID record with your professional information and link to your other identifiers (such as Scopus or ResearcherID or LinkedIn).
- 3 USE YOUR ORCID ID** Include your ORCID identifier on your Webpage, when you submit publications, apply for grants, and in any research workflow to ensure you get credit for your work.

MEMBERS MAKE ORCID POSSIBLE!

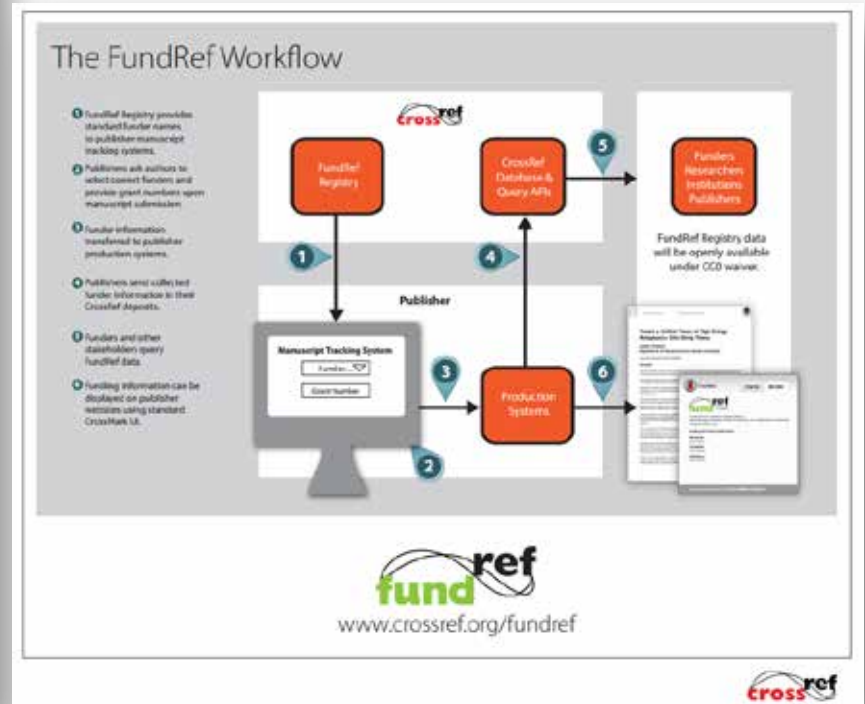
ORCID is a non-profit organization supported by a global community of organizational members, including research organizations, publishers, funders, professional associations, and other stakeholders in the research ecosystem.

Curious about who our members are? [See our complete list of member organizations.](#)

ORCID® is a registered trademark and the ORCID logo and ID icon are trademarks of ORCID, Inc. Other trademarks may be the property of their respective owners.

Contact Us | Privacy Policy | Terms of Use | Open Source Project

<http://orcid.org/>



http://www.crossref.org/08downloads/handouts/FundRef_Workflow.pdf

研究を革新した道具の評価

[PDF] Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry

K Tanaka, H Waki, Y Ido, S Akita... - ... mass spectrometry, 1988 - masspec.scripps.edu

... and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of-flight Mass Spectrometry

Koichi Tanaka¹, Hiroaki Waki ... on this method of sample preparation should be promising, enabling the detection of high ... 7. T. Yoshida, K. Tanaka, Y. Ido, S. Akita and Y. Yoshida, Mass ...

Cited by 2765 Related articles All 4 versions Cite Save

[CITATION] Detection of high mass molecules by laser desorption time-of-flight mass spectrometry

K Tanaka, Y Ido, S Akita, Y Yoshida, T Yoshida - ... Symposium on Mass Spectrometry, 1987

Cited by 67 Related articles Cite Save

Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of-flight Mass Spectrometry

Koichi Tanaka¹, Hiroaki Waki, Yutaka Ido, Satoshi Akita, Yoshikazu Yoshida and Tamio Yoshida

Shimadzu Corporation, Nishinokyo-Kuwabaracho, Nakagyo-ku, Kyoto 604, Japan

SPONSOR REFEREE: T. Matsuo, Osaka University, Osaka, Japan

Hitherto, ²⁵²Cf plasma desorption mass spectrometry (PDMS) has been used to study peptides and proteins in the molecular weight range from 1 kDa to 35 kDa.^{1,2} Fast atom bombardment mass spectrometry (FABMS)

Two TOF systems were constructed. The first system utilized a digital wave memory and accumulation circuits. This system could accumulate the spectrum data of 8 K words within 1 ms. In the first place, a "one shot"

Tanaka, Koichi, et al. "Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry." *Rapid communications in mass spectrometry* 2.8 (1988): 151-153.

「巨人の肩」賞

- **なんでも引用**の世界では、「巨人の肩」に値する研究を特定しやすくなるか？
- 最先端研究が「巨人の肩」になるとは限らない。**肩に乗るだけの研究**は対象外！
- 論文賞と「巨人の肩」賞は基準が異なる。引用を通して**多様な貢献**に光を当てる。
- 「巨人の肩」賞にふさわしい**研究（データ）基盤**とはどんなものか？

巨人の肩のわかりやすい図



株式会社稲葉製作所ウェブサイトより

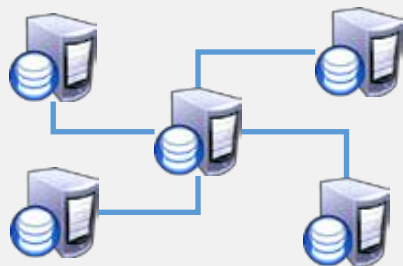
- 100人乗っても、1万人乗っても大丈夫な「巨人の肩」を作る人が偉いという世界観。

オープンサイエンス国内諸活動の連携

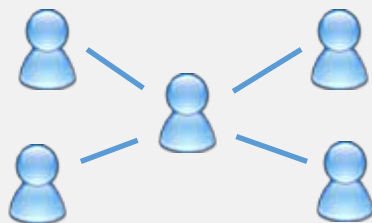
研究基盤の2つの要素

研究基盤(Research Infrastructure : RI)

データ基盤(DI)



コミュニティ基盤
(CI)



コミュニティ基盤
(Community
Infrastructure: CI)

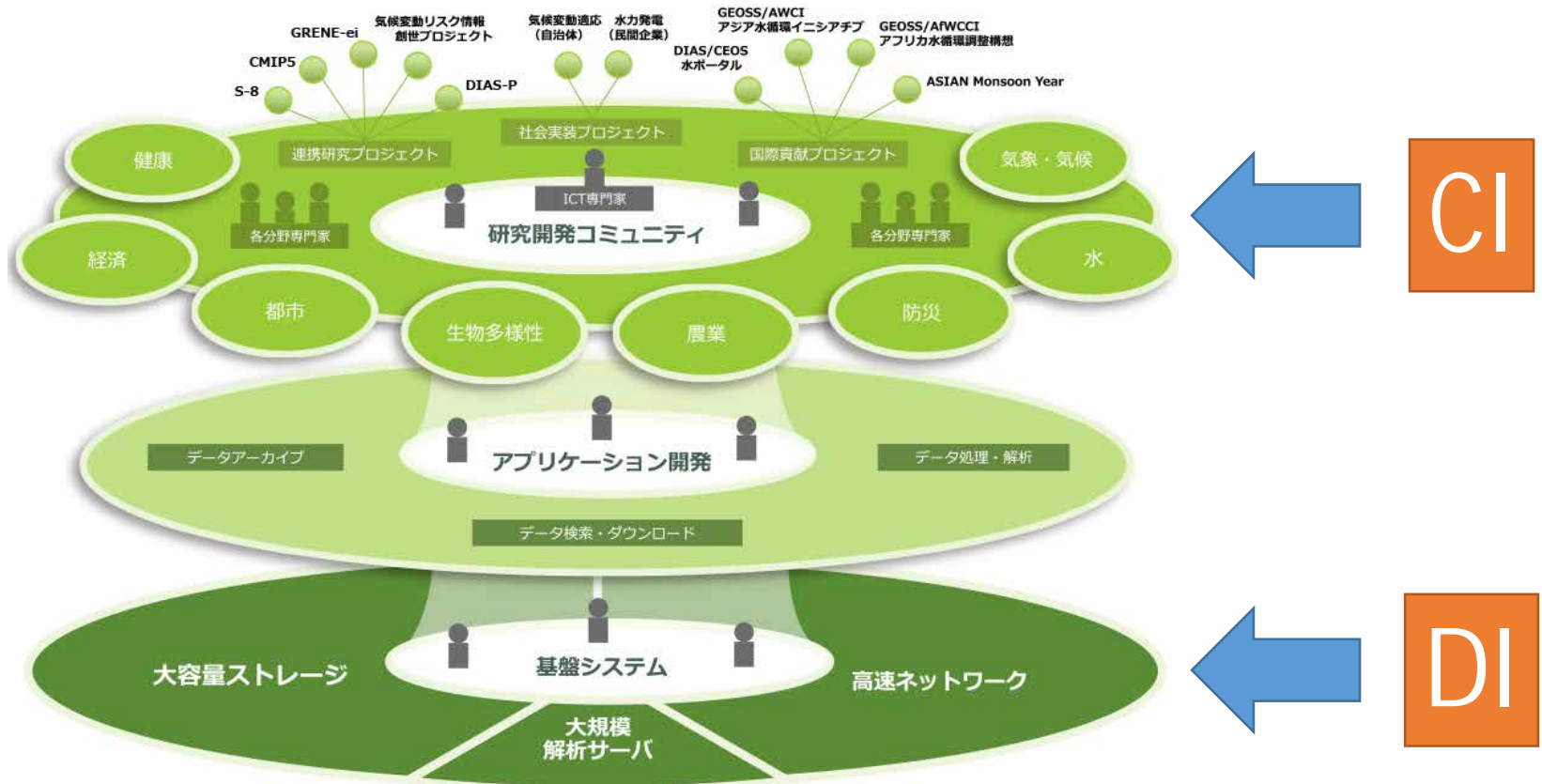
意見交換、要望、
コミュニケーション
のための環境

**データ基盤 (Data Infrastructure:
DI)**

研究データの集積、保存、管理

RI = DI + CI

DIAS = DI + CI



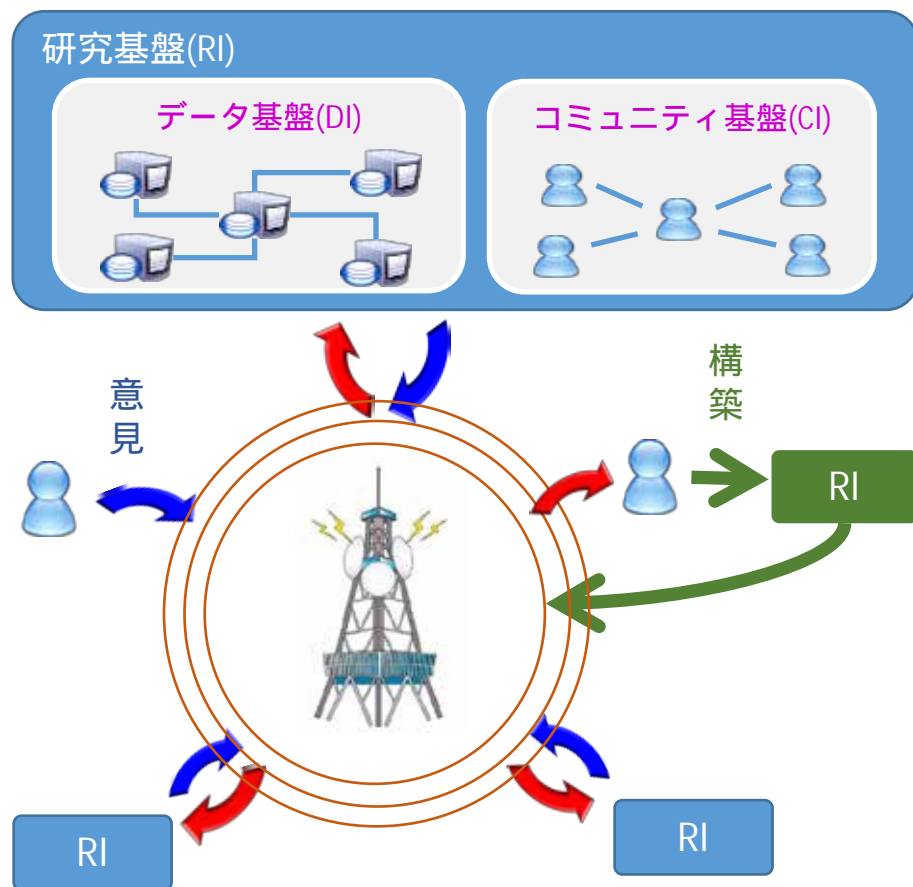
<http://www.diasjp.net/about/system/>

日本国内の認識不足

- オープンサイエンスに関して、報告書を除き、日本語のまとまった情報源がない。
- 研究データに関して、「データの現場」からの意見が可視化されていない。
- 研究基盤（Research Infrastructure）の世界的な発展から取り残されている。
- **オープンサイエンス・研究基盤に関する情報共有サイトが必要ではないか！**

「情報の塔」構想

- DIASの経験を踏まえて、研究データや研究基盤、オープンサイエンスに関する情報を集約するキュレーションサイト。
- 小野雅史氏（東京大学）を中心に検討中。
- 参考サイト：カレントアウェアネス（国立国会図書館）



図作成：小野雅史氏（東京大学）

情報共有サイトの内容（仮）

- オープンサイエンスに関するイベントカレンダー + 報告記事。
- 研究基盤やデータ共有に関する、国内や海外の動向の紹介。
- 研究者・技術者からの、研究基盤の現場感覚を反映した情報発信 / 意見集約。
- 国際的な研究基盤の連携や、日本国内の新たな研究基盤の構築支援。

国内諸活動の役割分担？

- **DIASオープンサイエンス勉強会** – 研究基盤の運用や情報共有サイト等の活動
- **JaLC** – DOIシステムを中心に日本版RDAのような取り組み等の活動
- **WDS + IUGONET** –サイエンス分野を中心にイベント開催等の活動
- **SPARC Japan + NII** – 情報学や図書館等のチャネルを通じた普及等の活動
- **地球研コアプロジェクトFS** – 市民科学や機関単位のアーカイブ等を通じた活動

オープンサイエンスの コミュニティ

DIAS

JaLC

研究データコミュニティ

研究者・学会・図書館・研究行政

まとめ

- オープンサイエンスとは、よりよい学術研究を目指す「夢」の収束点である。
- データサイテーションの「夢」は、データ活動の貢献が正当に評価されること。
- 国内諸活動の間で役割分担し、情報共有サイトを通して共通認識を深めていく。
- オープン化 = 善というよりは、**オープン化を梃子に新しい可能性を追究**したい。

関連ウェブサイト

- DIASプロジェクト
 - <http://www.diasjp.net/>
 - <https://dias.ex.nii.ac.jp/>
- 個人ページ
 - <http://agora.ex.nii.ac.jp/~kitamoto/>
 - <http://researchmap.jp/kitamoto/>