

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/>

Self Introduction



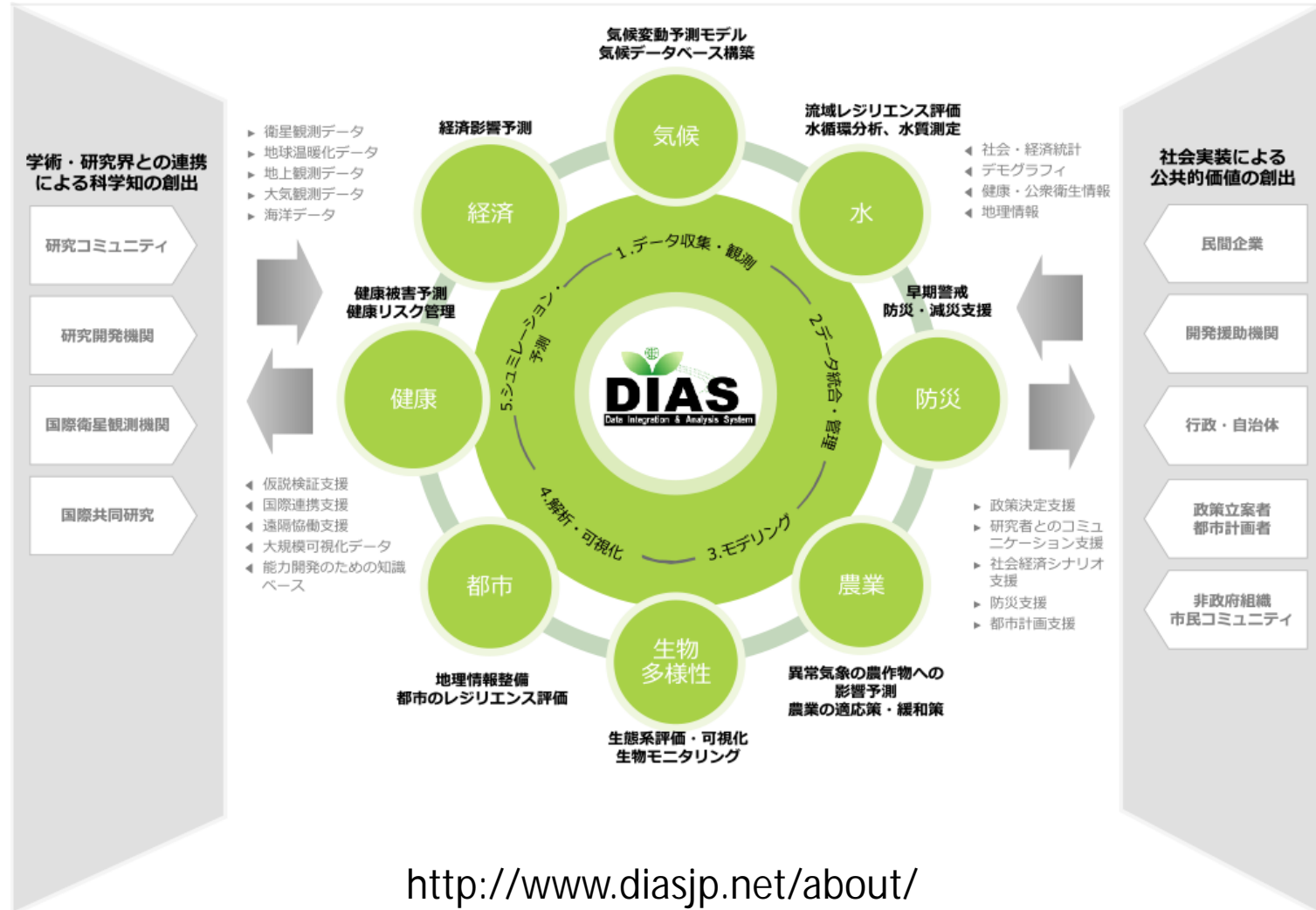
- Research on **creating values from data** at the forefront of informatics (boundary between other domains).
- Search and analysis on **visual data and geographic data**.
- Collaboration with scientists in other domains, especially **earth science or humanities**.

DIAS-P project



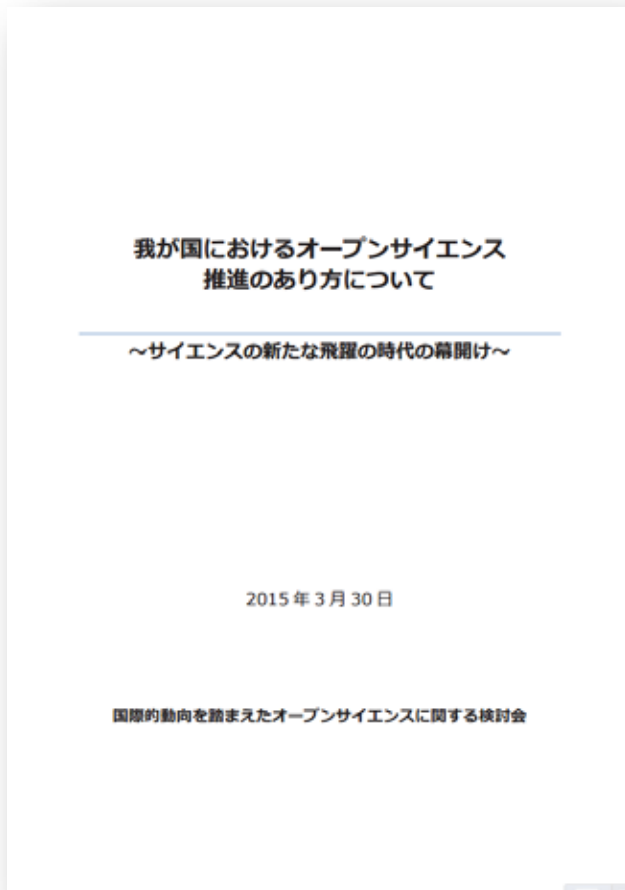
- Project duration: FY 2011 – FY 2015.
- One of MEXT projects on climate change adaptation strategy.
- I have been working on innovating the DIAS system for wider applicability and better utility.

Overview of DIAS-P



Convergence to Open Science

What is Open Science?



- **Open science** is a concept that includes **open access** and **open data**, leading to innovation through the adoption of openness on research results, knowledge and data.
- Improve access not only from academia but also from **industry and society** to research results by **funding from public sectors**.

http://www8.cao.go.jp/cstp/sonota/openscience/150330_openscience_1.pdf



オープンサイエンス革命
マイケル・ニールセン (著)
紀伊國屋書店, 2013
右の引用はp.340.

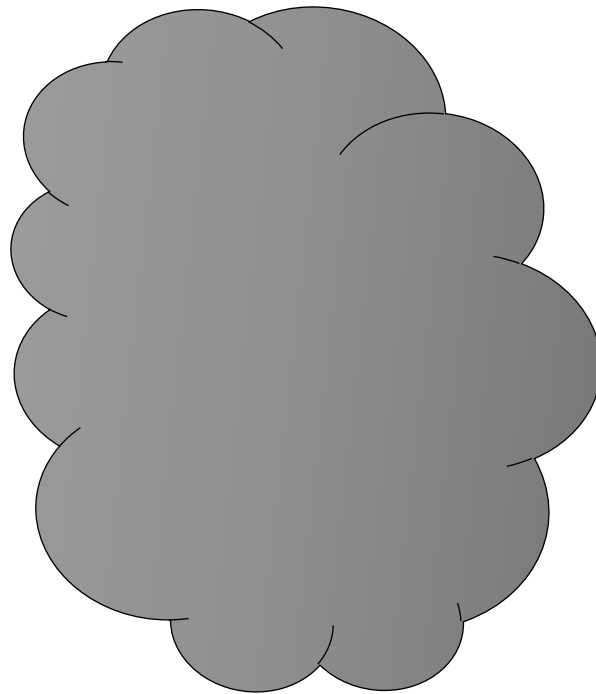
- Science has its own **problems** to slow down a movement to openness.
- A tool such as CC does not offer an exact solution on **credit issues** that scientists are evaluated by the publication of papers.
- Open science can learn from general activities toward openness, but **it also needs new ideas.**

My Definition

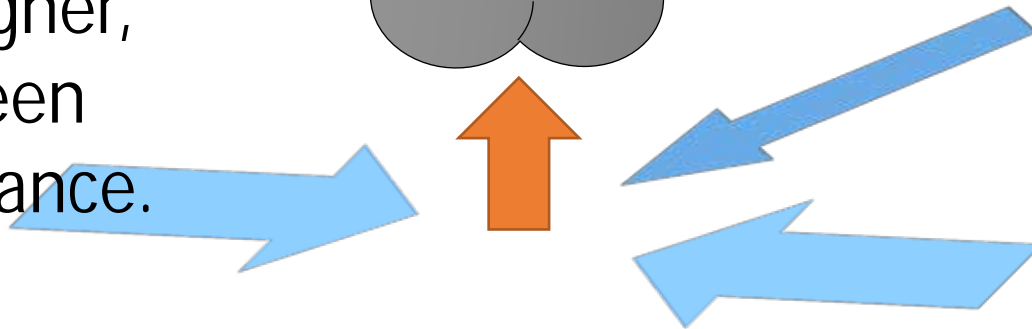
Open Science is the **convergence of dreams** by people who are not satisfied with the current practice of science, and see a **possibility of revolution by leveraging the concept of openness.**

Convergence to Open Science

A cumulonimbus is growing higher at a point of convergence caused by many movements. As growing higher, it can be seen from a distance.



Everyone has a different dream, but their hopes seem to be converging to the same point.



Different Dreams in Open Science

Different Dreams

1. Open Data
2. Open Access
3. Reproducible Science
4. Citizen Science
5. Data Citation
6. Open Innovation,

(1) Open Data

Problems

- The value of data is not maximized in the society even if it is supported by public funding.
- Data will not be reused by others if scientists never release their data.

Dreams

- Using **open data license** for effective data sharing and reuse.
- More data sharing and reuse may be realized by **“open research data license”** that considers the needs of research data.

What is Open Data?

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

Key Features of Openness

- Availability and access
- Reuse and redistribution
- Universal participation

Open data has two dimensions, namely **“value (utility)”** and **“transparency”**.

Why Open Data?

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

(2) Open Access

Problems

- **Serials crisis**: too expensive journals to continue purchase.
- **Dominating companies**: Only a few commercial companies can survive large-scale investment on electronic journals.

Dreams

- **Green OA** (Embargo +repository) or **Gold OA** (article processing charge).
- Recommendations or rules on open access through the power of **funding agencies**.

Movement in Funding Agencies

The screenshot shows the NSF website with the following content:

- Header:** National Science Foundation, WHERE DISCOVERIES BEGIN. Navigation menu: FUNDING, AWARDS, DISCOVERIES, NEWS, PUBLICATIONS, STATISTICS, ABOUT NSF, FASTLANE.
- Left Sidebar:** Office of Budget, Finance and Award Management (BFA) logo and navigation links: DIAS Home, CAAR Branch, Policy Office, Systems Office, View DIAS Staff, Search DIAS Staff, BFA Organization (Office of Budget, Finance, & Award Management, Budget Division, Division of Acquisition and Cooperative Support, Division of Financial Management, Division of Grants & Agreements, Division of Institution & Award Support).
- Main Content:**
 - Section:** Dissemination and Sharing of Research Results
 - Sub-section:** NSF Data Sharing Policy
 - Text:** Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. See [Award & Administration Guide \(AAG\) Chapter VI.D.4](#).
 - Section:** NSF Data Management Plan Requirements
 - Text:** Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See [Grant Proposal Guide \(GPG\) Chapter II.C.2.1](#) for full policy implementation.
 - Section:** Requirements, by Directorate, Office, Division, Program, or other NSF Unit
 - Text:** Links to data management requirements and plans relevant to specific Directorates, Offices, Divisions, Programs, or other NSF units, are provided below. If guidance specific to the program is not provided, then the requirements established in [Grant Proposal Guide, Chapter II.C.2.1](#) apply.
 - Text:** Please note that if a specific program solicitation provides guidance on preparation of data management plans, such guidance must be followed.
 - List:**
 - Biological Sciences Directorate (BIO)
 - [Directorate-wide Guidance](#)
 - Computer & Information Sciences & Engineering (CISE)
 - [Directorate-wide Guidance](#)
 - Education & Human Resources Directorate (EHR)
 - [Directorate-wide Guidance](#)
 - Engineering Directorate (ENG)
 - [Directorate-wide Guidance](#)
 - Geosciences Directorate (GEO)
 - [Directorate-wide Guidance](#)

NSF Data Management Plan

The screenshot shows the Bill & Melinda Gates Foundation website with the following content:

- Header:** BILL & MELINDA GATES Foundation. Search bar and language selector (EN).
- Section:** BILL & MELINDA GATES FOUNDATION OPEN ACCESS POLICY
- Text:** The Bill & Melinda Gates Foundation is committed to information sharing and transparency. We believe that published research resulting from our funding should be promptly and broadly disseminated. We have adopted an Open Access policy that enables the unrestricted access and reuse of all peer-reviewed published research funded, in whole or in part, by the foundation, including any underlying data sets.
- Text:** As of January 1, 2015 our Open Access policy will be effective for all new agreements. During a two-year transition period, publishers will be permitted to apply up to a 12-month embargo period on the accessibility of the publication and its underlying data sets. This embargo period will no longer be allowed after January 1, 2017.
- Text:** Our Open Access policy contains the following elements:
- List:**
 1. **Publications Are Discoverable and Accessible Online.** Publication will be deposited in a specified repository(s) with proper tagging of metadata.
 2. **Publication Will Be On "Open Access" Terms.** All publications shall be published under the [Creative Commons Attribution 4.0 Generic License \(CC BY 4.0\)](#) or an equivalent license. This will permit all users of the publication to copy and redistribute the material in any medium or format and transform and build upon the material, including for any purpose (including commercial) without further permission or fees being required.
 3. **Foundation Will Pay Necessary Fees.** The foundation would pay reasonable fees required by a publisher to effect publication on these terms.
 4. **Publications Will Be Accessible and Open Immediately.** All publications shall be available immediately upon their publication, without any embargo period. An embargo period is the period during which the publisher will require a subscription or the payment of a fee to gain access to the publication. We are, however, providing a transition period of up to two years from the effective date of the policy (or until January 1, 2017). During the transition period, the foundation will allow publications in journals that provide up to a 12-month embargo period.
 5. **Data Underlying Published Research Results Will Be Accessible and Open Immediately.** The foundation will require that data underlying the published research results be immediately accessible and open. This too is subject to the transition period and a

Open Access Policy

BILL & MELINDA GATES FOUNDATION OPEN ACCESS POLICY (JAN. 2015)

- Publications Are Discoverable and Accessible Online.
- Publication Will Be On “Open Access” Terms.
- Foundation Will Pay Necessary Fees.
- Publications Will Be Accessible and Open Immediately.
- Data Underlying Published Research Results Will Be Accessible and Open Immediately.

(3) Reproducible Science

Problems

- Many research results are **not reproducible**, and credibility of research is in danger.
- **Closed software and research environment** hampers verification of research results.

Dreams

- Open research data enables verification by anyone.
- Possibility of verification against misconduct leads to **transparency of research**.

(4) Citizen Science

Problems

- Scientific data collection could be improved through the participation of citizens.
- Accountability and social responsibility is limited without participation of citizen to scientific processes.

Dreams

- **Participatory platform** can lead to problem solutions with the power of citizen.
- **Trans-disciplinary science** : co-creation of values through collaboration between scientists and citizen.

Participatory Monitoring



Bombus terrestris was introduced from Europe in 1991, and designated as invasive species in 2006.

- We continued participatory monitoring in Hokkaido since 2006.
- Participants capture the invasive bee and report the situation to scientists.

Collaboration with Prof. Izumi Washitani (University of Tokyo) as part of DIAS project.

Seiyou Status

<http://www.seiyoubusters.com/seiyou/>

セイヨウ情勢：市民参加による外来種（セイヨウオオマルハナバチ）をモニタリングと対策のためのリアルタイム情報共有サイト。最終更新日：2013年02月06日

Seiyou Status / Partnership Monitoring for Invasive Alien Species

セイヨウ情勢

「セイヨウ情勢」ウェブサイトは、市民参加型の監視活動（モニタリング）を通じてみなさんから寄せられたセイヨウオオマルハナバチの捕獲・目撃情報を公開しています。セイヨウオオマルハナバチの侵入状況と対策のいまを知るために、ぜひご利用ください。【もっと詳しく...】【English Version...】

2012年度の活動を開始します。今年度もご協力をよろしくお願い申し上げます。昨年度の活動結果は2011年活動報告書をご覧ください。

最新情報はこちら！
セイヨウ情勢ブログ

参加者からのおたよりと運営者からのひとこと

2012年の活動成果

2012年の最新状況

| | |
|----------|----------|
| 確認市町村数 | 60 |
| 活動参加者数 | 294 |
| セイヨウ同撃数 | 50,399 頭 |
| 在来種目撃数 | 3,930 頭 |
| セイヨウ捕獲数 | 43,508 頭 |
| セイヨウ女王バチ | 5,335 頭 |
| 巣 | 12 個 |

活動報告書

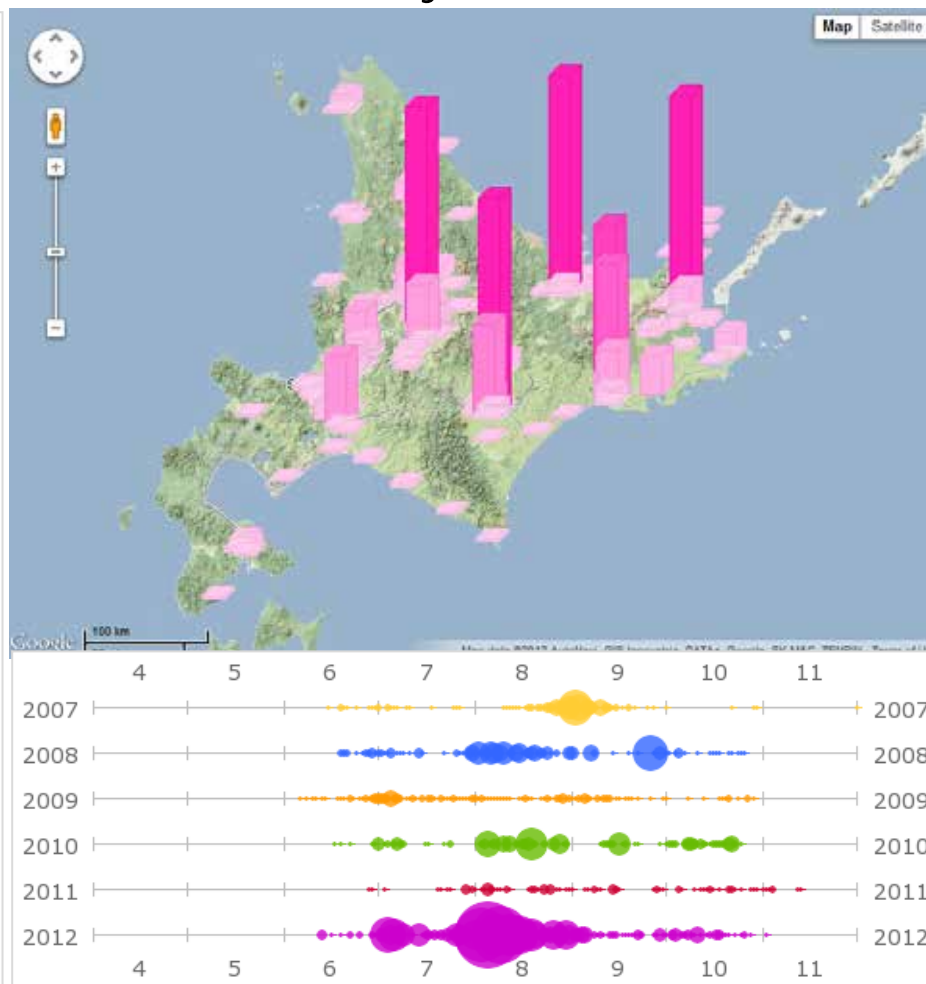
- 2011
- 2010
- 2009
- 2008
- 2007
- 2006

年ごとの情報

- 2012年

セイヨウマップ 在来種マップ

条件を指定してリスト・地図・グラフを表示する



(5) Data Citation

Problems

- Construction and operation of data infrastructure does not receive proper appreciation.
- Data-related research needs new framework for evaluation toward sustainable career.

Dreams

- **Data papers and data citations** are a new approach for evaluating data-related research.
- **Scientists are familiar with the practice of citation** and we already have infrastructure for citations.

Example of Data Journals

The screenshot shows the homepage of the journal 'Scientific Data'. The header includes the journal title 'SCIENTIFIC DATA' and a search bar. Below the header, there are navigation links: Home, Archive, About, For Authors, For Referees, Data Policies, and Collections. The main content area features a 'Featured Data Descriptor' with a waveform image and a text description: 'The Icelandic 16-electrode electrohysterogram database'. Below this, there is a 'Latest Research' section with a grid of article previews, each including a title, authors, and a date. On the right side, there is an 'About Scientific Data' section, a 'Submit manuscript' button, and an 'Open for submissions' banner for the 'npj Systems Biology and Applications' journal. At the bottom, there are 'Announcements' and 'Scientific Data Update' sections.

Scientific Data (Nature publishing group)

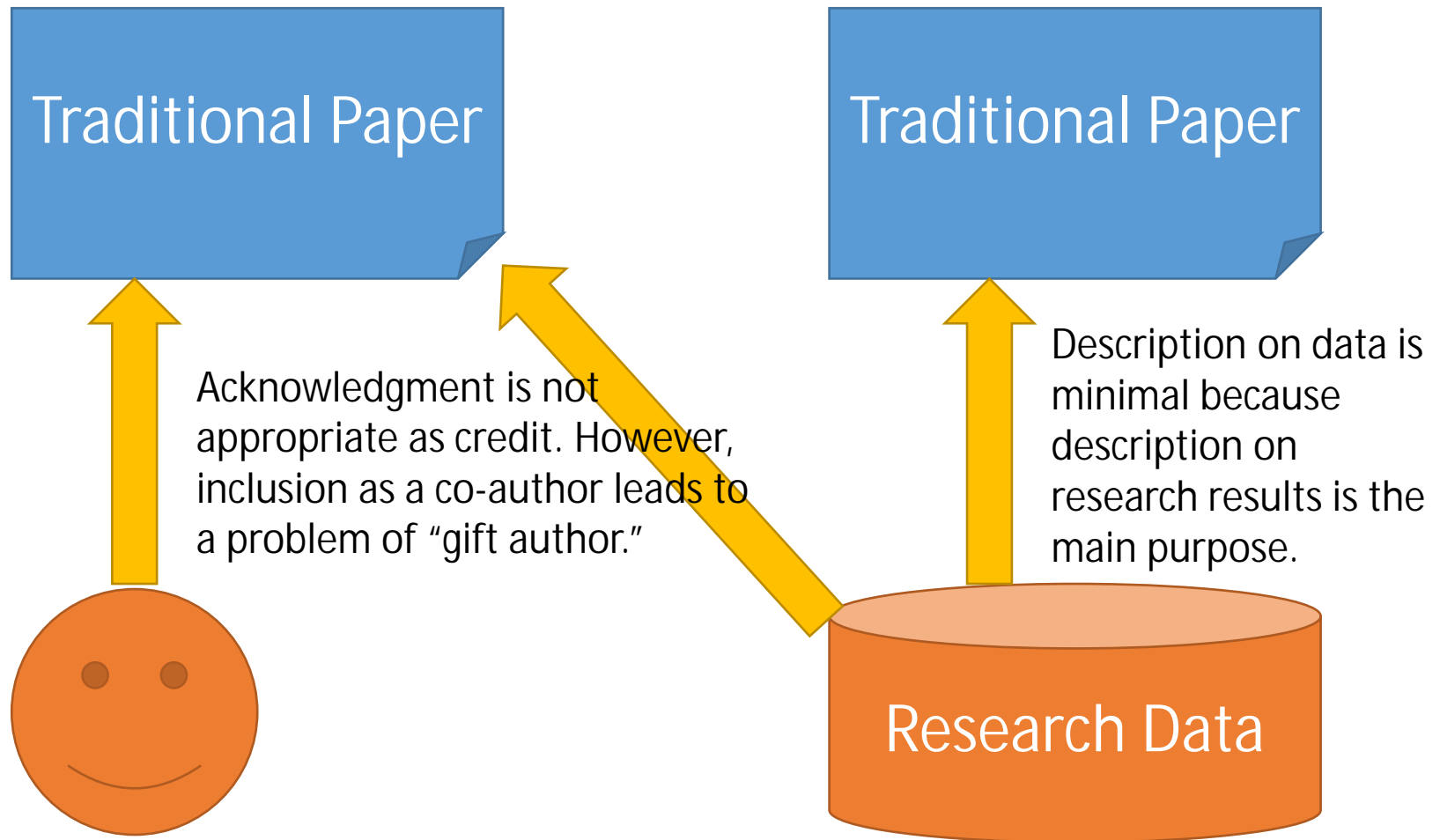
The cover of the journal 'Earth System Science Data' is shown. The top half has a blue background with the journal title 'Earth System Science Data' in white. Below the title, it says 'The Data Publishing Journal'. The middle section features a photograph of several pens resting on a grid, with the numbers '20' and '119' visible. On the right side of the image, the URL 'www.earth-system-science-data.net' is written vertically. The bottom section has a dark blue background with the text 'Volume 1 • Issue 1 • 2009'. At the very bottom, the Copernicus Publications logo and name are displayed, along with the tagline 'The Innovative Open Access Publisher'.

Earth System Science Data (Copernicus)

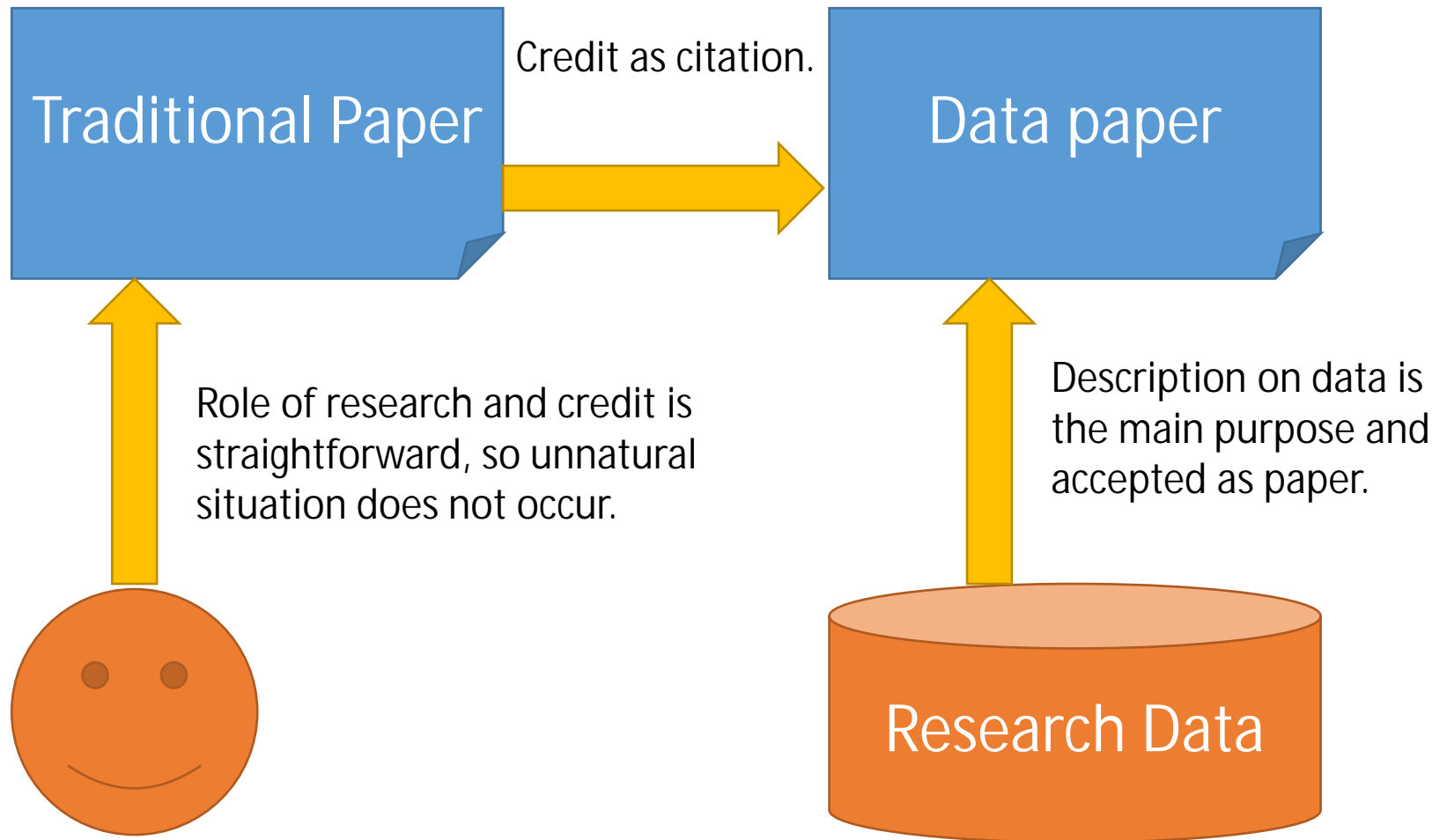
Review of Data Journals

- Scientific Data has the following criteria.
 1. Experimental Rigor and Technical Data Quality
 2. Completeness of the Description
 3. Integrity of the Data Files and Repository Record
- **Importance or novelty of discovery from the dataset is not evaluated.** Interpretation of data should not be included.
- Data papers focus useful description for data interpretation and reuse, such as quality control, and the rigor of data generation process.

Traditional Publication



Publication with Data Papers



Fostering Communities for Shared Perception

DIAS Open Science Seminar

- DIAS members (NII, Univ. Tokyo, Kyoto Univ. JAMSTEC, NIES) gathered as a working group for the JaLC experimental project.
- ML started on Oct. 2014, and the seminar started on Jan. 2015. **We had 10 meetings**, and discussed various themes such as identifiers, metadata and open science.
- We contributed the JaLC project by requests on improvement and documentation.

Topics Discussed in the Seminar

- Progress report on JaLC experimental project.
- Review of data repositories in the world.
- DOI assignment rules and version control.
- Clarification on agenda about data citation.
- What is dynamic citation?
- Requirement for recommended repositories.
- Information sharing system on open science.
- Cultural difference in different communities.

Shared Perception

- **I** had little knowledge, even on DOI, so we started to explore related information.
- **We** started to have a global picture by sharing information on various topics.
- **The seminar** expanded by welcoming DIAS members from NIES to have a wider view.
- **JaLC experimental project** consists of people with different motivation on open science.

Community of Open Science

DIAS

JaLC

Community on research data

Scientists, Societies, Libraries, Government

My Dream

“Shoulders of Giants”



ウェブ全体から検索 日本語のページを検索

巨人の肩の上に立つ

Articles (include patents) Case law

Stand on the shoulders of giants

- Academic research can have a far-reaching view by standing on the achievement of past research.
- A scientist who made “shoulders of giants” is great. However, are they receiving proper appreciation?

Revolutionary Tool

[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.

ID Systems for Research

Linking researchers and funding agencies

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!

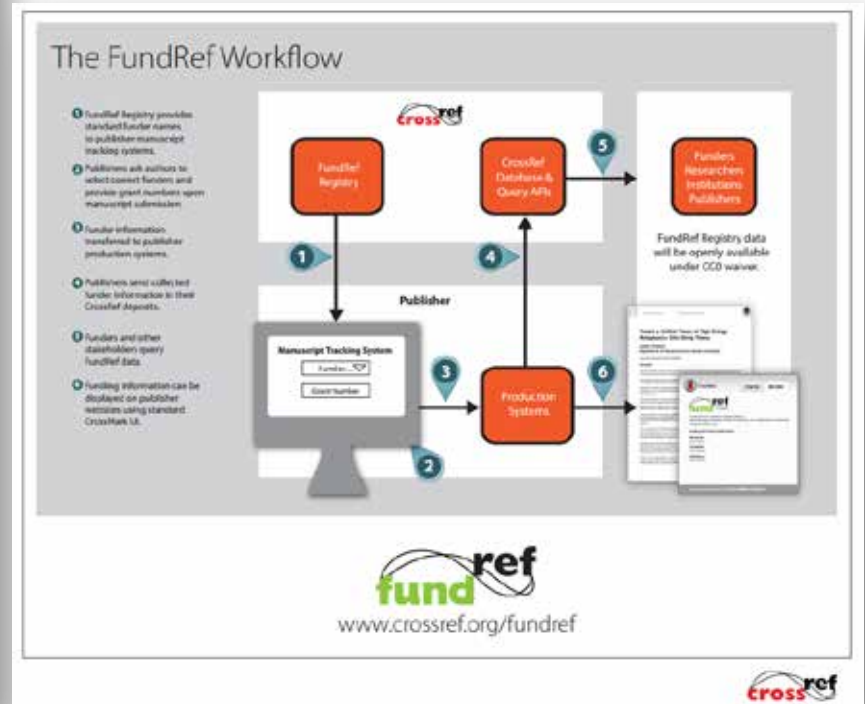
ORCID is a non-profit organization supported by a global community of organizational members, including research organizations, publishers, funders, professional associations, and other stakeholders in the research ecosystem.

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

Software Citation

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', 'Writing 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 make code citable by archiving it on Zenodo. A 'Pro Tip' section offers advice for researchers. The 'Choose your repository' section explains that repositories are the most 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. The header is blue with the Zenodo logo and the tagline 'Research. Shared.'. The main content area is white with a blue sidebar on the right. The main text area shows the title 'Analysis of Cartesian Genetic Programming's Evolutionary Mechanisms' by Brian W. Goldman. Below the title is a 'Preview' section showing a file tree for the repository 'briangoldman/Analysis-of-CGP's-Mechanisms-v0.0.0'. The file tree lists various files and their sizes. The 'Files' section at the bottom shows the file 'Analysis of CGP's Mechanisms 0.0.0.zip' with a size of 53.5 kB and a date of 11 May 2015. The right sidebar contains a 'Publication date' section, a 'DOI' section, a 'Published in' section, a 'Related publications and datasets' section, a 'Collections' section, a 'License (for files)' section, and a 'New to Zenodo?' section.

<https://zenodo.org/>

Total Evaluation of Research

- **Citation of everything**: if all research results are citable, the existing **citation platform** can be used for the quantification of researcher's contribution and grant's contribution.
- Citation platform that maintains linked research outcome may lead to **more complex research metrics** that considers research contribution not in traditional papers.

Giant Shoulder Award

- Data citation, or **citation of everything**, is useful to identify **giant research contribution**?
- **State-of-the-art research** is not always “giant shoulders” because some scientists just take advantage of it.
- What is **data infrastructure** that deserves “giant shoulder award”? Every academic society could establish its own metric.

Visualization of Giant Shoulders?



株式会社稲葉製作所ウェブサイトより

- Giant shoulder is stable even if 100 people, or 10,000 people, are relying on it.

Conclusion

Summary

- Open Science is the convergence of dreams to realize better scientific research.
- “Open=Good” is not the reason; openness is for leveraging the revolution of science.
- People see different hopes in “openness.”
- Community about open science is necessary for the continuous development of shared perception from different dreams.

Related Websites

- **DIAS Project**

- <http://www.diasjp.net/>
- <https://dias.ex.nii.ac.jp/>

- **Personal Webpage**

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