人工知能がもたらす学術コミュニケーションの変容

Nobuko Miyairi

http://orcid.org/0000-0002-3229-5662

@NobukoMiyairi

PIAJ
Pharmaceutical Information Association of Japan

2019.6.21 PIAJ 2019年度総会 講演会

自己紹介

1991-1999	東京外国語大学附属図書館
2000-2002	ハワイ大学マノア校図書館情報学修士
2003-2012	Thomson Reuters（現Clarivate Analytics）
2012-2015	Nature Publishing Group（現Springer Nature)
2015-2018.4	ORCID, Inc. (APAC Regional Director)
2018.5-	学術情報コンサルタント（フリーランス）
	Paper Digest（Strategic Adviser)
2019.1-	World Data System (Communications Officer)
2019.4-	筑波大学（非常勤講師）

はじめに

- 本講演の内容は、「実務家」として図書館および学術情報流通業界で働いた個人的経験をもとに考察を重ねたものです。
- 本講演で述べる意見や見解は個人的なものであって、現在または過去に所属した／している組織の意向を反映したものではありません。
- 一部の内容、特に人工知能に関する部分については、極めて限定的な知識と経験をベースとしています。
- 本講演で言及するサービス等は、決して網羅的なものではなく、あくまでも選択的に事例として紹介しています。

構成

- （個人的な）問題意識
- 現在進行中のサイエンス2.0、あるいはオープンサイエンス
- 前提としてのサイエンス1.0
- 人工知能とは何か
- 人工知能を使った学術情報サービスの事例
- 人工知能がもたらす学術コミュニケーションの変容（私論）
- ディスカッション

（個人的な）問題意識

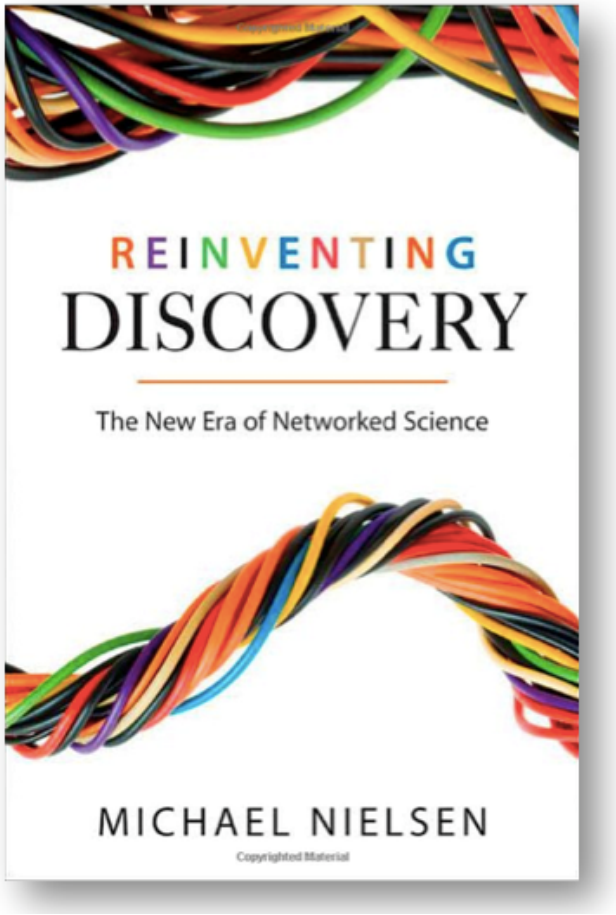
1. 「学術コミュニケーション」と「技術革新」
    - 活版印刷
    - インターネット
    - 人工知能
2. 「量」と「質」
    - 計量書誌学
    - 研究評価
    - 科学技術予測
3. 学術コミュニケーションにおける「アクター」
    - 生産者
    - 利用者
    - 仲介者

（個人的な）問題意識

1. 「学術コミュニケーション」と「技術革新」
    - 活版印刷
    - インターネット
    - 人工知能
2. 「量」と「質」
    - 計量書誌学
    - 研究評価
    - 科学技術予測
3. 学術コミュニケーションにおける「アクター」
    - 生産者
    - 利用者
    - 仲介者

ネットワークが可能にする新しいサイエンス

- 共通言語・認識
- 細分化された専門性
- 細分化された貢献
- モジュラー化を通したスケールアップ
- 「デザインされた」セレンディピティ



Nielsen, M. (2012). *Reinventing discovery: The new era of networked science.* Princeton, NJ: Princeton University Press.

情報管理 JOHO KANRI 2014 vol.57 no.2 5 May
Journal of Information Processing and Management
http://johokanri.jp/

オープンサイエンスと科学データの可能性

Open science and scientific data

宮入 暢子[1]

MIYAIRI Nobuko[1]

1 ネイチャー・パブリッシング・
E-mail: n.miyairi@nature.com
1 Nature Publishing Group (Chiyo

原稿受理 (2014-03-11)
情報管理 57(2), 080-089, doi: 10.1241/joh

宮入, 暢子. "オープンサイエンスと科学データの可能性." *情報管理* 57, no. 2 (2014): 80-89.

著者抄録

Galaxy Zoo や eBird に代表さ
民が直接貢献している。「開か
科学研究の基本理念である (1) 先駆性の確保、(2) 科学の集約化、(3) 第三者による正当性の担保、(4) 著者による説明責任の確立、といった基礎を築いた。サイエンス 2.0 の到来によって、プレプリント、オープンピアレビュー、オープンデータリポジトリ、科学のソーシャル化によるネットワークを介したイノベーションなど、学術コミュニケーションの多様化が促進された。これまで論文とその引用という形でしか計ることのできなかった研究インパクトに、オンライン上の注目度を定量化するオルトメトリクスが加わり、各国政府の研究データのオープン化の方針が進む中、科学データ流通を促進するための情報基盤の確立は急務である。

キーワード

オープンサイエンス, シチズンサイエンス, オープンアクセス, オープンデータ, データ再利用, 科学史, サイエンス2.0, オルトメトリクス

1. はじめに

本稿では，17世紀後半に成立した学術誌とピアレビューを中心とする科学情報の流通基盤が，シチズンサイエンスやサイエンス2.0など現在のオープンサイエンスの枠組みの中で，どのような局面を迎えているのかを概観する。オープンなコラボレーションと批判精神は，古くから学術コミュニケーションに内在するものであり，オープンデータと科学のソーシャル化がそれらを促進したことは，必ずしも「新しい」現象ではないと筆者は考える。むしろ科学技術とネットワークの発達によって「ようやく」可能になったものであるという視点を与えることで，次々と登場する新たな学術情報サービスに必然性を見いだし，将来を展望する一助となれば幸いである。

2. 市民が支える科学

オープンサイエンスの提唱者マイケル・ニールセンは，その著書Reinventing Discovery（邦題：オープンサイエンス革命）[1] の中で，オンラインネットワークやデジタルツールを活用して科学研究の過程で得られたデータをオープンに共有し，より効率よく発展させようという試みを多数紹介している。ここで何度も繰り返し登場するのが，Galaxy Zoo[注1] である。Galaxy Zooは，スローン・デジタル・スカイ・サーベイ（Sloan Digital Sky Survey: SDSS）[注2] によって集められた何百万もの銀河系の画像を分類・整理することを目的として2007年7月から始まったWebサイトで，運用開始初日のうちにアップロードされた7万点余りの画像が，アマチュア天文家の手により分類された。当初，ボランティアの参加者たちは，銀河が楕円形なのか，渦巻きが時計回りなのか反時計回りなのか，という単純な，しかしコンピューターには不可能な判定を目視により行っていた。そのうち何人かが，分類作業を進めるうちに緑色のこれまで見たことのない銀河の画像を発見し，Webサイト上に設けられたWikiを利用してディスカッションを始めた。ついには「グリーンピース銀河（Green Peas galaxy）」と名付けられた新しいタイプの銀河について，これらのボランティアを共著者とする研究論文が出版されるまでに至った。

鳥類の観測データを収集・提供するオンラインデータベースeBird[注3] には，2002年以来これまでに9万人以上のボランティアが参加し，アップロードされるデータ量は毎年40パーセント程度増加しているという。**図1**はイヌワシ（Golden Eagle）の北米における分布をeBird上の機能を使って示したものである。彼らがアップロードしたデータは，このようにインタラクティブなマップとして表示したり，研究用データとしてダウンロードしたりすることができ，鳥類の研究はもとより，自然環境や生物多様性の保全など，さまざまな方面で利用されている。

Galaxy ZooもeBirdも，いわゆるシチズンサイエンスの代表例といえるが，研究者を本職としないアマチュアによる科学への貢献はインターネットの登場以前からすでにあった。米国オーデュボン協会が主催するChristmas Bird Count[注4] は，1900年に開始された当初27名のボランティアによるささやかな試みであったが，113回目にあたる2013年の調査は7万人以上が参加する大きなプロジェクトとなった。こうしたボランティアが長年積み上げたデータによって，北米に生息する20種ほどの鳥類が，過去40年間に平均で100マイルほど北に生活範囲を移動したことが明らかになり（**図2**），地球温暖化や気象調査の背景データとして使われている[注5]。

3. ソーシャルな科学

Polymath Project[注6] もまた，ニールセンがしばしば取り上げる興味深い事例である。英国の数学者でフィールズ賞受賞者でもあるティモシー・ガウアーズが2009年1月に自身のブログ上で発した問い（Is massively collaborative mathematics possible?）は，そもそもニールセンのブログポスト[注7] に触発されたものだった。ガウアーズは「もし数学者がたくさん集まって，効率よく思考を進めれば，効率よく問題

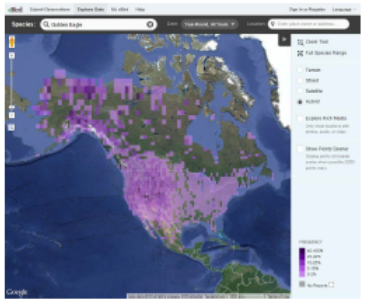


図1 イヌワシ（Golden Eagle）の分布

CITIZEN SCIENCE

Data on Wings

A modest effort to enlist amateur bird-watchers in the cause of ornithology wound up producing a fire hose of data and helping rewrite the rules of science

By Hillary Rosner

IN BRIEF

Automated-sensor networks monitor much of our environment, but some data collection in the digital age still requires the efforts and close analyses of phalanxes of context-sensitive human beings who can help solve problems of scale.

A field called citizen science, which involves public participation in research, marshals laypeople's observations, often by way of high-tech consumer devices and machines.

Based at the Cornell Lab of Ornithology, in collaboration with the National Audubon Society, eBird is one of the most mature such efforts. It and its ilk have yielded academic-caliber results in astronomy, computer science and public health, while giving skilled amateurs more opportunities to contribute.

Rosner, H. Data on Wings. *Scientific American* 308, 68-73 (14 January 2013).

Changes in the Winter Ranges of Birds

Audubon

www.christmasbirdcount.org

PROJECTS ABOUT GET INVOLVED TALK BUILD A PROJECT NEWS SIGN IN REGISTER

Galaxy Zoo

ABOUT CLASSIFY TALK COLLECT

GALAXY ZOO STATISTICS

70% Complete

21,603	1,781,265	51,868	36,369
Volunteers	Classifications	Subjects	Completed Subjects

WORDS FROM THE RESEARCHER

"In the decade the project has been running, Galaxy Zoo volunteers have helped understand the Universe and made spectacular discoveries. We hope you'll join us for the next stage of the adventure."

ABOUT GALAXY ZOO

To understand how galaxies formed we need your help to classify them according to their shapes. If you're quick, you may even be the first person to see the galaxies you're asked to classify.

Look at telescope images of distant galaxies.

Explore the sky. What will you find?

https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/

Astrophysics > Cosmology and Extragalactic Astrophysics

Galaxy Zoo Green Peas: Discovery of A Class of Compact Extremely Star-Forming Galaxies

Carolin N. Cardamone, Kevin Schawinski, Marc Sarzi, Steven P. Bamford, Nicola Bennert, C. M. Urry, Chris Lintott, William C. Keel, John Parejko, Robert C. Nichol, Daniel Thomas, Dan Andreescu, Phil Murray, M. Jordan Raddick, Anze Slosar, Alex Szalay, Jan VandenBerg

(Submitted on 23 Jul 2009)

We investigate a class of rapidly growing emission line galaxies, known as "Green Peas", first noted by volunteers in the Galaxy Zoo project because of their peculiar bright green colour and small size, unresolved in SDSS imaging. Their appearance is due to very strong optical emission lines, namely [O III] 5007 A, with an unusually large equivalent width of up to ~1000 A. We discuss a well-defined sample of 251 colour-selected objects, most of which are strongly star forming, although there are some AGN interlopers including 8 newly discovered narrow Line Seyfert 1 galaxies. The star-forming Peas are low mass galaxies (M~10^8.5 - 10^10 M_sun) with high star formation rates (~10 M_sun/yr), low metallicities (log[O/H] + 12 ~ 8.7) and low reddening (E(B-V) < 0.25) and they reside in low density environments. They have some of the highest specific star formation rates (up to ~10^{-8} yr^{-1}) seen in the local Universe, yielding doubling times for their stellar mass of hundreds of Myrs. The few star-forming Peas with HST imaging appear to have several clumps of bright star-forming regions and low surface density features that may indicate recent or ongoing mergers. The Peas are similar in size, mass, luminosity and metallicity to Luminous Blue Compact Galaxies. They are also similar to high redshift UV-luminous galaxies, e.g., Lyman-break galaxies and Lyman-alpha emitters, and therefore provide a local laboratory with which to study the extreme star formation processes that occur in high-redshift galaxies. Studying starbursting galaxies as a function of redshift is essential to understanding the build up of stellar mass in the Universe.

Comments: 18 pages, including 13 figures and 4 tables, Accepted for publication by MNRAS
Subjects: **Cosmology and Extragalactic Astrophysics (astro-ph.CO)**; Galaxy Astrophysics (astro-ph.GA)
Cite as: arXiv:0907.4155 **[astro-ph.CO]**
(or arXiv:0907.4155v1 **[astro-ph.CO]** for this version)

https://arxiv.org/abs/0907.4155

SCIENCE 2.0

BY M. MITCHELL WALDROP

Is posting raw results online, for all to see, a great tool or a great risk?

The first generation of World Wide Web capabilities rapidly transformed retailing and information search. More recent attributes such as blogging, tagging and social networking, dubbed Web 2.0, have just as quickly expanded people's ability not just to consume online information but to publish it, edit it and collaborate about it—forcing such old-line institutions as journalism, marketing and even politicking to adopt whole new ways of thinking and operating.

Science could be next. A small but growing number of researchers (and not just the younger ones) have begun to carry out their work via the wide-open tools of Web 2.0. And although their efforts are still too scattered to be called a movement—yet—their experiences to date suggest that this kind of Web-based "Science 2.0" is not only more collegial than traditional science but considerably more productive.

"Science happens not just because of people doing experiments but because they're discussing those experiments," explains Christopher Surridge, managing editor of the Web-based journal Public Library of Science On-Line Edition (www.plosone.org). Critiquing, suggesting, sharing ideas and data—this communication is the heart of science, the most powerful tool ever invented for correcting errors, building on colleagues' work and fashioning new knowledge. Although the classic peer-reviewed paper is important, says Surridge, who publishes a lot of them, "they're effectively just snapshots of what the authors have done and thought at this moment in time. They are not collaborative beyond that, except for rudimentary mechanisms such as citations and letters to the editor."

Web 2.0 technologies open up a much richer dialogue, says Bill Hooker, a postdoctoral cancer researcher at the Shriners Hospital for Children in Portland, Ore., and author of a three-part survey on open-science efforts that appeared at 3 Quarks Daily (www.3quarksdaily.com), where a group of bloggers write about science and culture. "To me, opening up my lab notebook means giving people a window into what I'm doing every day," Hooker says. "That's an immense leap forward in clarity. In a paper, I can see

JON KRAUSE

KEY CONCEPTS

- Science 2.0 generally refers to new practices of scientists who post raw experimental results, nascent theories, claims of discovery and draft papers on the Web for others to see and comment on.
- Proponents say these "open access" practices make scientific progress more collaborative and therefore more productive.
- Critics say scientists who put preliminary findings online risk having others copy or exploit the work to gain credit or even patents.
- Despite pros and cons, Science 2.0 sites are beginning to proliferate; one notable example is the OpenWetWare project started by biological engineers at the Massachusetts Institute of Technology.

—The Editors

Waldrop, M. Science 2.0. *Scientific American* 298, 68-73 (May 2008).

SCIENTIFIC AMERICAN™

Hello, nmiyairi (sign out)

Search ScientificAmerican.com

Subscribe | News & Features | Topics | Blogs | Videos & Podcasts | Education | Ci

More Science :: Edit This :: January 9, 2008 :: 139 Comments :: Email :: Print

Science 2.0: Great New Tool, or Great Risk?

Wikis, blogs and other collaborative web technologies could usher in a new era of science. Or not.

By M. Mitchell Waldrop

Welcome to a Scientific American experiment in "networked journalism," in which readers—you—get to collaborate with the author to give a story its final form.

The article, below, is a particularly apt candidate for such an experiment: it's my feature story on "Science 2.0," which describes how researchers are beginning to harness wikis, blogs and other Web 2.0 technologies as a potentially transformative way of doing science. The draft article appears here, several months in advance of its print publication, and we are inviting you to comment on it. Your inputs will influence the article's content, reporting, perhaps even its point of view.

So consider yourself invited. Please share your thoughts about the promise and peril of Science 2.0.—just post your inputs in the Comment section below. To help get you started, here are some questions to mull over:

What is *Edit This*?

Edit This is your chance to take part in the creation of a feature article for an upcoming issue of *Scientific American*.

Help us shape the content by adding your questions and comments below.

Pin it

Image:

KEY CONCEPTS

- Science 2.0 generally refers to new practices of scientists who post raw experimental results, nascent theories, claims of discovery and draft papers on the Web for others to see and comment on.
- Proponents say these "open access" practices make scientific progress more collaborative and therefore more productive.
- Critics say scientists who put preliminary findings online risk having others copy or exploit the work to gain credit or even patents.
- Despite pros and cons, Science 2.0 sites are beginning to proliferate; one notable example is the OpenWetWare project started by biological engineers at the Massachusetts Institute of Technology.

先駆性

正当性

オープン

RISK OF BEING SCOOPED

Dr. Monica: My first thought was, no way am I making my scientific ruminations public property. I've learned over the years that this is a sure way to have those ideas appear in someone else's work! However, many practical and useful applications come to mind.

Funklord: The issue is not that someone is just going to replicate your work and claim credit. The issue is, What if they are able to reach the eureka moment faster than you are?

NO PROXY FOR PEER REVIEW

Darren: One of the big positives of the current journal system is peer review: Science 2.0 needs a reputation-management system, a central database responsible for tracking the reputation of those participating in the online community.

wilbanks: Blogs and wikis are the digital equivalents of the hallway conversations at a conference or a lab meeting, but they are a long way from replacing journals. You don't get points for making a statement first in science unless you can prove that statement.

EXPANDED OPENNESS

Deadlyvices: Web 2.0 has fantastic potential to open up science to everyone, not just tenured academics. Perhaps if intelligent laypeople had a greater opportunity to contribute, there would be less public disaffection with science.

Richaa: One reason I left science after my Ph.D. was the isolationist culture. One scientist told me I had too many interests to be successful in physics research. I decided to take that as a compliment. I hope that the opening of science through Web 2.0 technologies will remove that culture and bring in valuable interdisciplinary thinking and collaboration.

効率性

説明責任

Current model	Emerging model
Research done privately; then submitted to journals; then peer-reviewed by gatekeepers in major journals; published	Research data shared during discovery stages; ideas shared; scientists collaborate; then findings are disseminated online
Scientific literature behind paywalls online	Scientific discoveries free online
Credit established by name attached to journal article.	How is credit established for contributors?
Data is private until publication	Data is shared before publication
Papers generally protected by copyright	Many different licenses possible: copyright, public domain, Creative Commons 3.0, etc.
Publishers raise funds by charging for access to content	Publishers seek alternative funding models
Journal article summaries available online after publication	Share methods, data, findings via blogs, social networking sites wikis, computer networking, Internet, video journals

http://en.wikipedia.org/wiki/Science_2.0

サイエンス1.0から2.0へ

サイエンス1.0	サイエンス2.0
研究は密かに行われ、結果としての論文をジャーナルに投稿→査読を経た後に出版	アイディアやデータをシェアして他の研究者と協働することが求められ、結果は即座にオンラインでシェア
出版社は購読モデルにもとづいて、読者に対して課金	出版モデルはよりオープンに、読者よりは著者に課金
出版のタイミングと掲載された著者名によって、貢献度や先駆性が認知される	貢献度や先駆性の認知は著者だけではない？
データは出版まで公開しない	データは出版前でも公開
論文は一般に著作権によって保護	著作権以外にも、クリエイティブ・コモンズなど様々なライセンス形態が存在
論文の副産物は、論文出版後にのみ公開	研究の方法、データ、結果などはブログ、SNS、ビデオなどを介して「ポスト」

http://en.wikipedia.org/wiki/Science_2.0 をもとに、宮入が日本語でまとめたもの

サイエンス1.0がもたらしたもの

先駆性の確保

- 「発見者」のみが享受できる「栄誉」
- ノーベル賞

科学の集約化

- 学術研究の「高騰」
- パトロンによる援助から国家予算へ

第三者による正当性の担保

- 科学アカデミーの成立
- ピア・レビュー

著者による説明責任の確立

- 文献引用、典拠

（個人的な）問題意識

1. 「学術コミュニケーション」と「技術革新」
   - 活版印刷
   - インターネット
   - 人工知能
2. 「量」と「質」
   - 計量書誌学
   - 研究評価
   - 科学技術予測
3. 学術コミュニケーションにおける「アクター」
   - 生産者
   - 利用者
   - 仲介者

ジャーナル数の増加

「科学計量学の父」と呼ばれるデレク・デ・ソラ・プライスは、学術ジャーナルの数は幾何級数的に増え、300を超えたあたりでレビュージャーナルが登場したことを指摘

1960年代前半に、2000年までにジャーナル数は100万タイトルを超えると予想

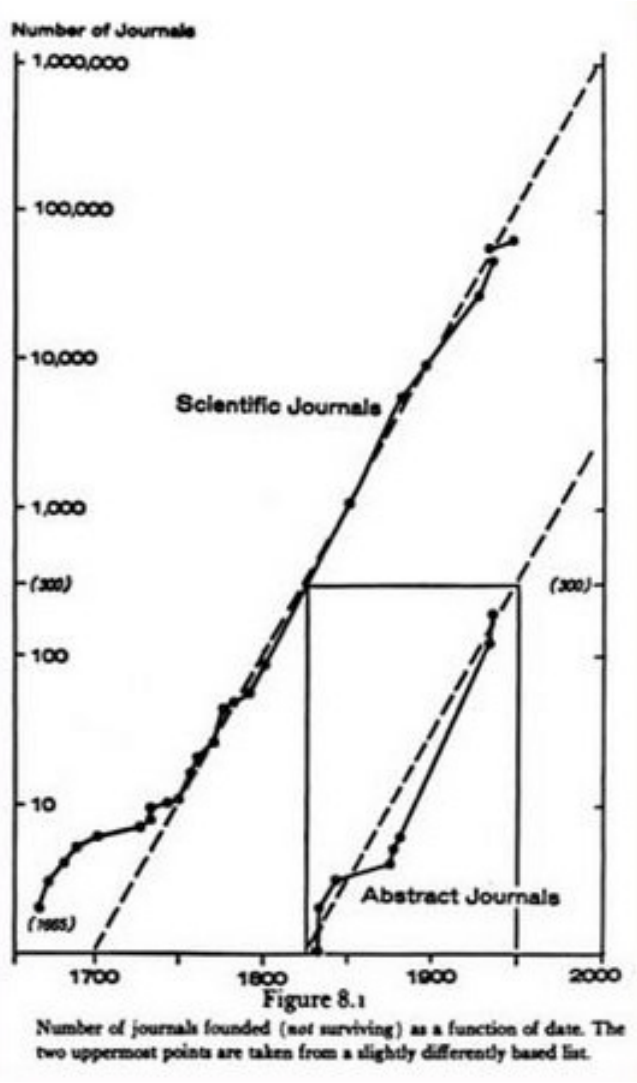
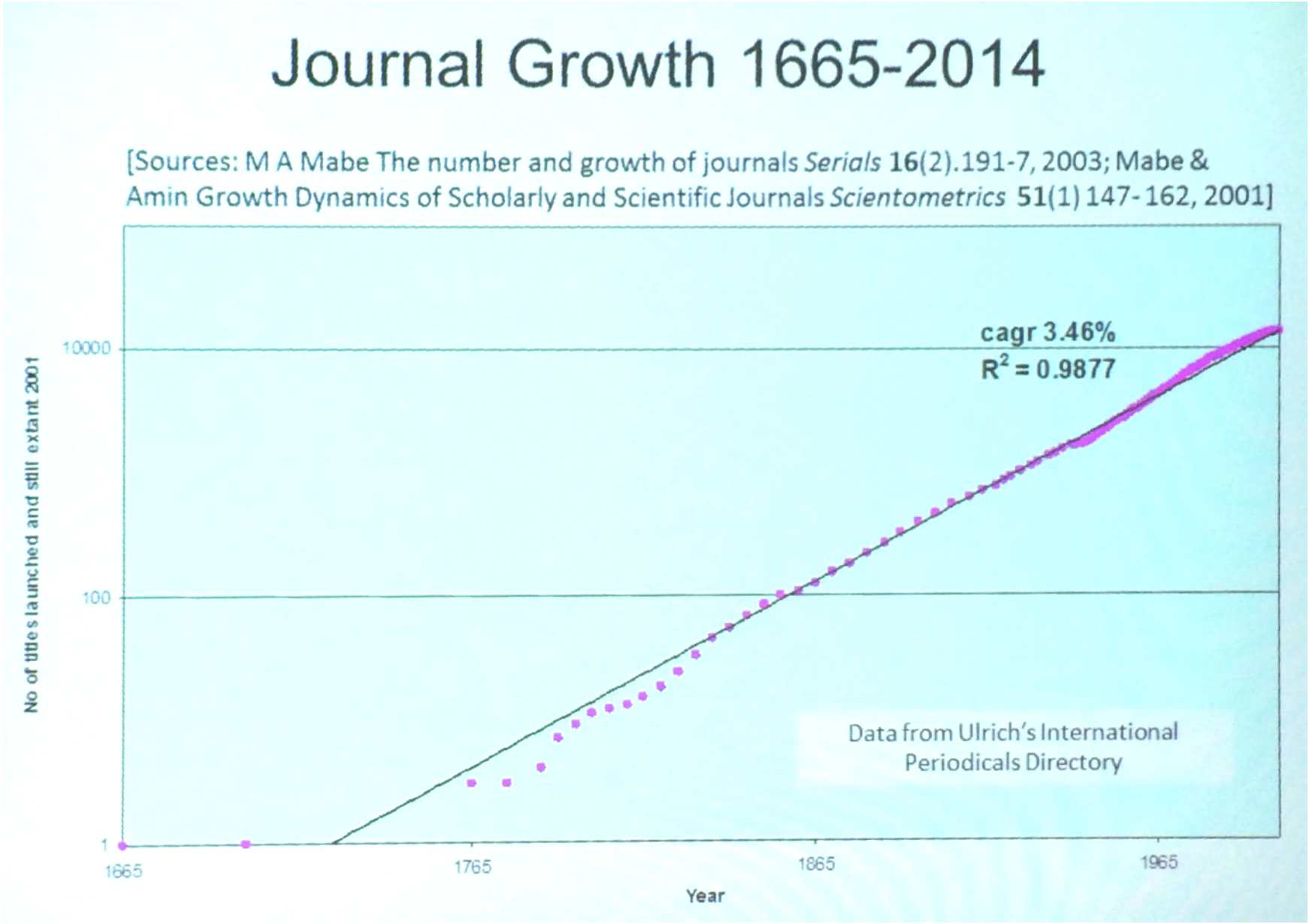


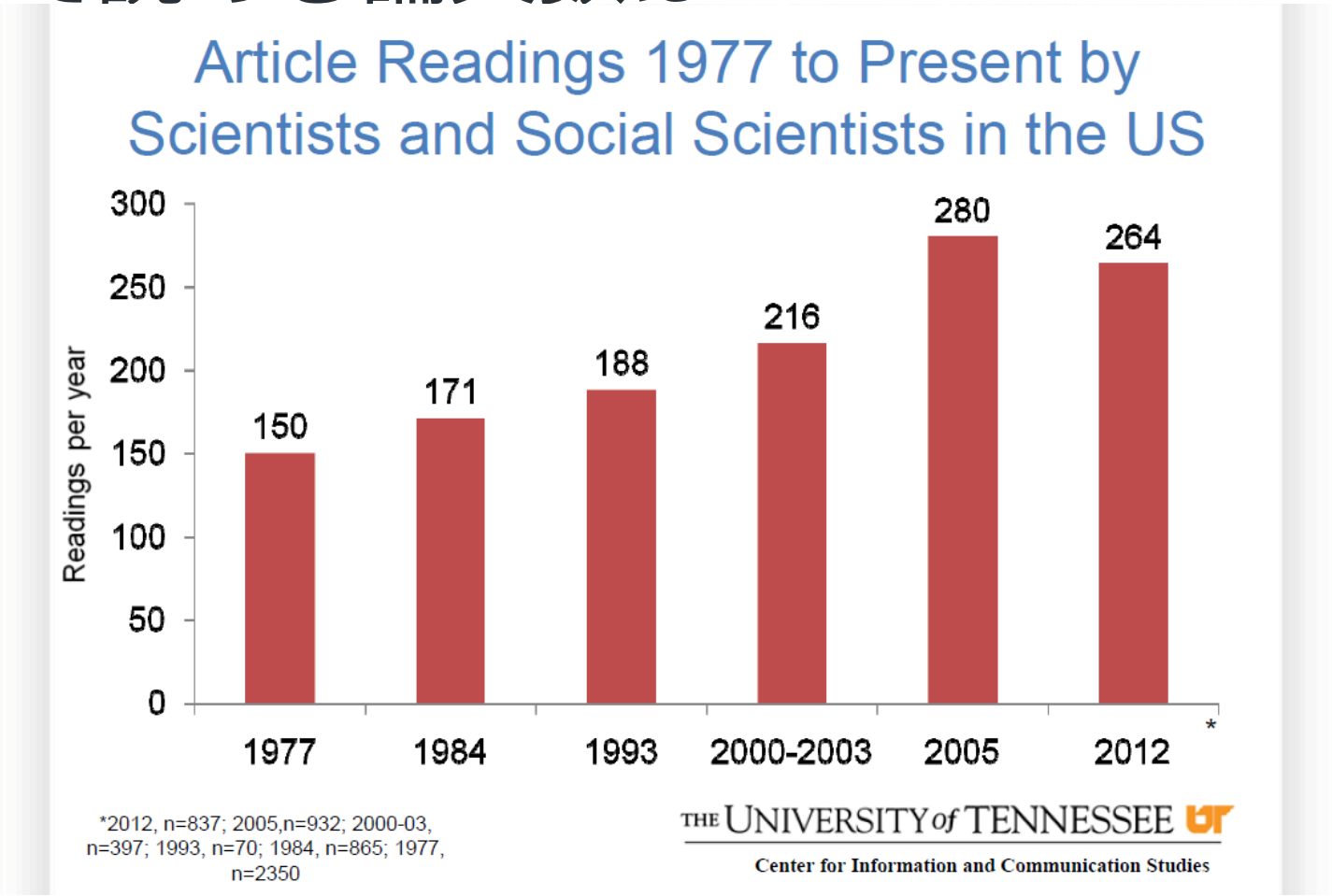
Figure 8.1
Number of journals founded (not surviving) as a function of date. The two uppermost points are taken from a slightly differently based list.

Price, Derek J. de Solla (1963). Little science, big science. New York: Columbia University Press.

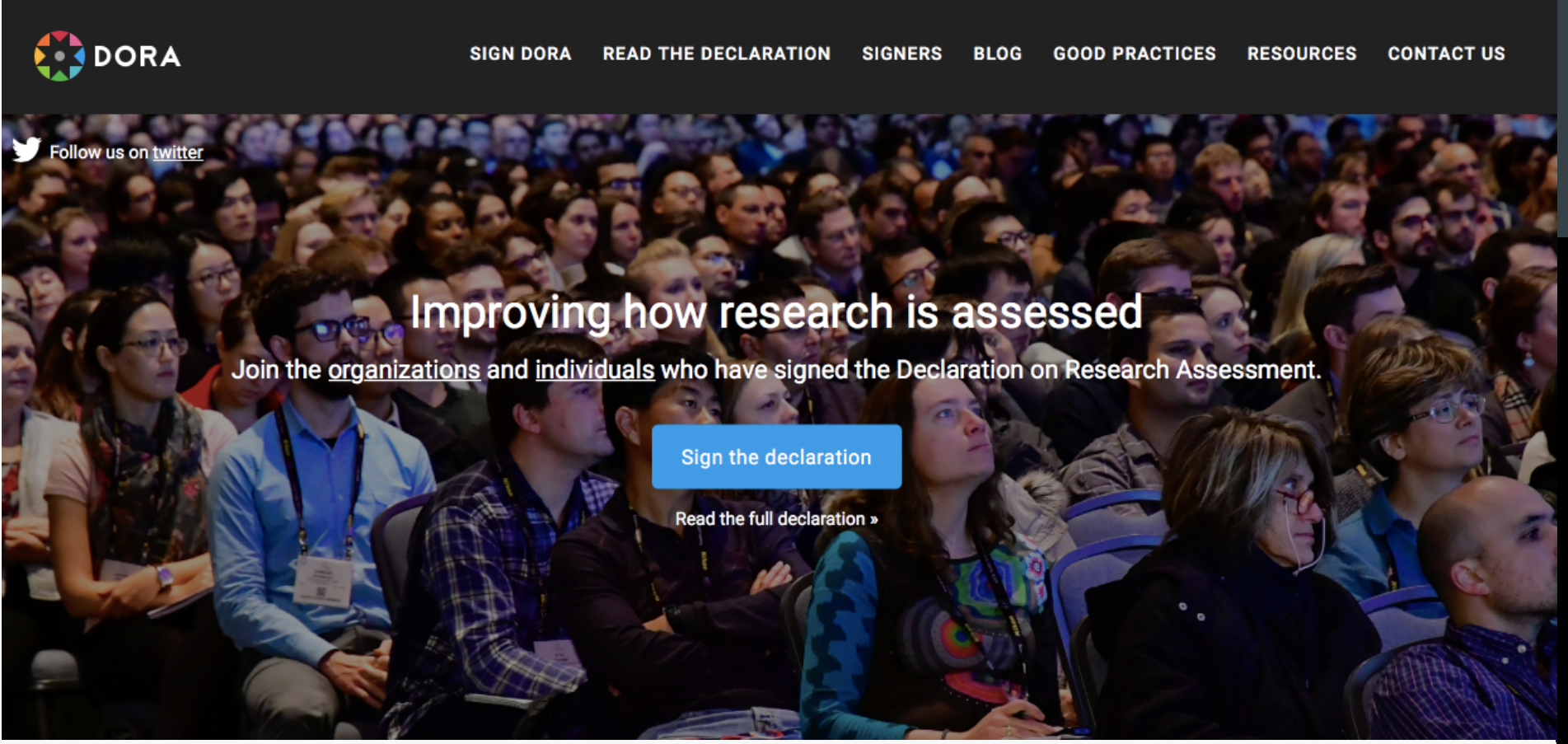


Mabe, M. Henry Oldenburg & the 350th Birthday of the Learned Journal. 2015. [presentation]; Photo by Miyairi, N.

1年で読める論文数は？



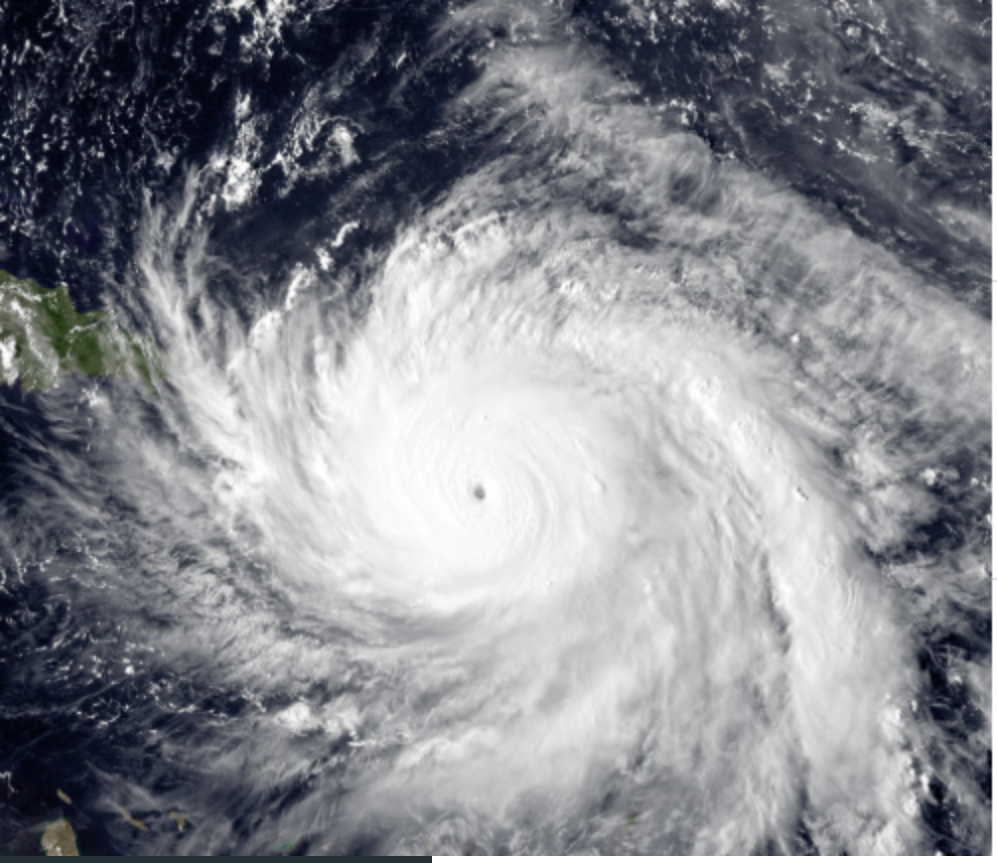
Tenopir, C. Altmetrics and Traditional Metrics: What Do Scholars Use to Judge Quality? Fiesole Collection Development Retreat, 2013.



About

The Declaration on Research Assessment (DORA) recognizes the need to improve the ways in which the outputs of scholarly research are evaluated. The declaration was developed in 2012 during the Annual Meeting of the American Society for Cell Biology in San Francisco. It is a worldwide initiative covering all scholarly disciplines and all key stakeholders including funders, publishers, professional societies, institutions, and researchers. We encourage all individuals and organizations who are interested in developing and promoting best practice in the assessment of scholarly research to sign DORA.

https://sfdora.org



FREE

#1 of 100

Mortality in Puerto Rico after Hurricane Maria

In the face of inadequate government data on mortality rates following one of the decade's biggest catastrophes, researchers released this new approach to determining the number of deaths caused by natural disasters.

Published in	New England Journal of Medicine
Date	July 2018
Subject area	Medical & Health Sciences

- **446** News stories
- **51** Blog posts
- **1** Policy document
- **16,847** Tweets
- **21** Facebook posts
- **5** Wikipedia mentions
- **12** Google+ posts
- **5** Reddit posts
- **1** Q&A site post

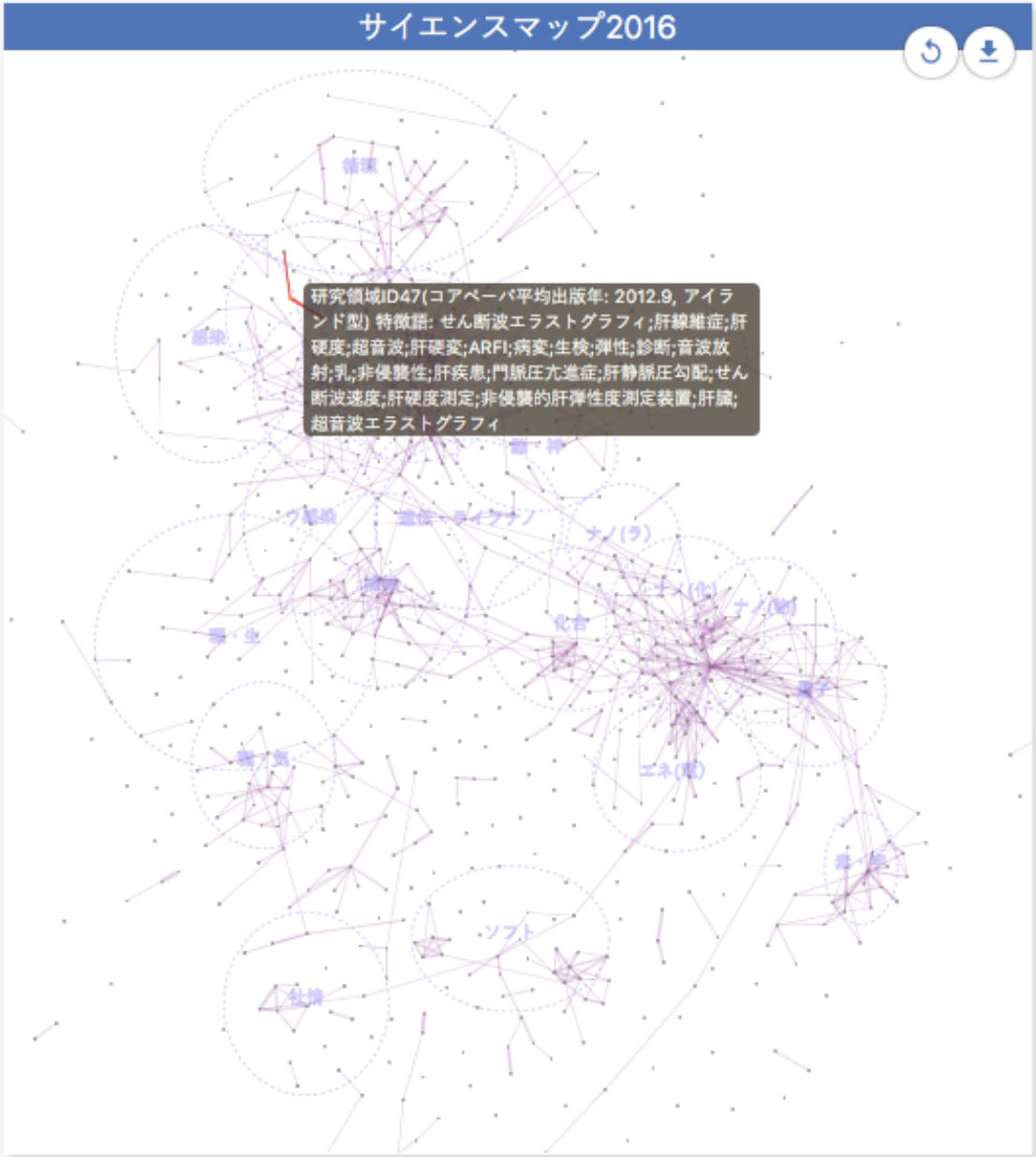
- **130** Readers on Mendeley
- **18** Citations on Dimensions

In the top 5% of all research outputs scored by Altmetric

One of the highest-scoring outputs from this source (#1 of 24,792)

High Attention Score compared to outputs of the same age (99th percentile)

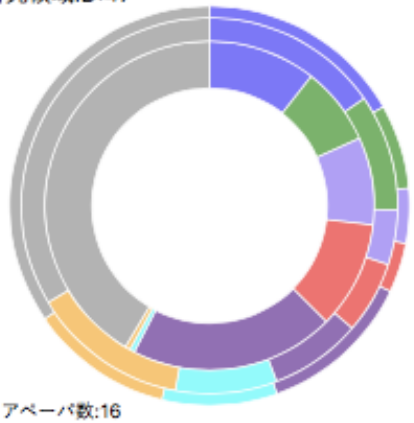
High Attention Score compared to outputs of the same age and source (99th percentile)



マウスドラッグ	マップの位置調整
マウスホイール	マップの拡大と縮小
灰色の円をマウスオーバー	特徴語(上位20程度)の表示
灰色の円をマウスクリック	主要国シェアとStreamの表示
Alt+マウスドラッグ	指定された範囲の特徴語のワードクラウド表示

主要国シェア(分数カウント)

研究領域ID:47



コアペーパ数:16
サイティングペーパ(Top10%)数:953
サイティングペーパ数:208

可視化対象の該当領域数

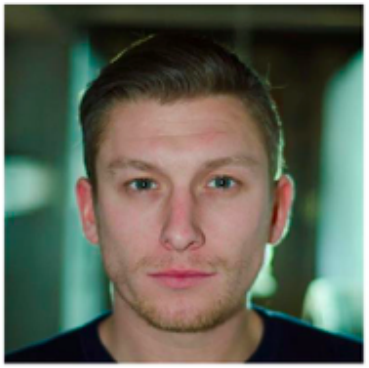
研究領域		--
コアペーパ	1件以上	0
サイティングペーパ(Top10%)	1件以上	0
サイティングペーパ	2件以上	0
特徴語を含む領域数 (検索語が複数ある場合、語によって色が異なります)	重複除く	0
該当する論文シェアの領域数		0

http://www.nistep.go.jp/wp/wp-content/uploads/ScienceMapWebEdition2016.html

（個人的な）問題意識

1. 「学術コミュニケーション」と「技術革新」
    - 活版印刷
    - インターネット
    - 人工知能
2. 「量」と「質」
    - 計量書誌学
    - 研究評価
    - 科学技術予測
3. 学術コミュニケーションにおける「アクター」
    - 生産者
    - 利用者
    - 仲介者

研究者スタートアップ



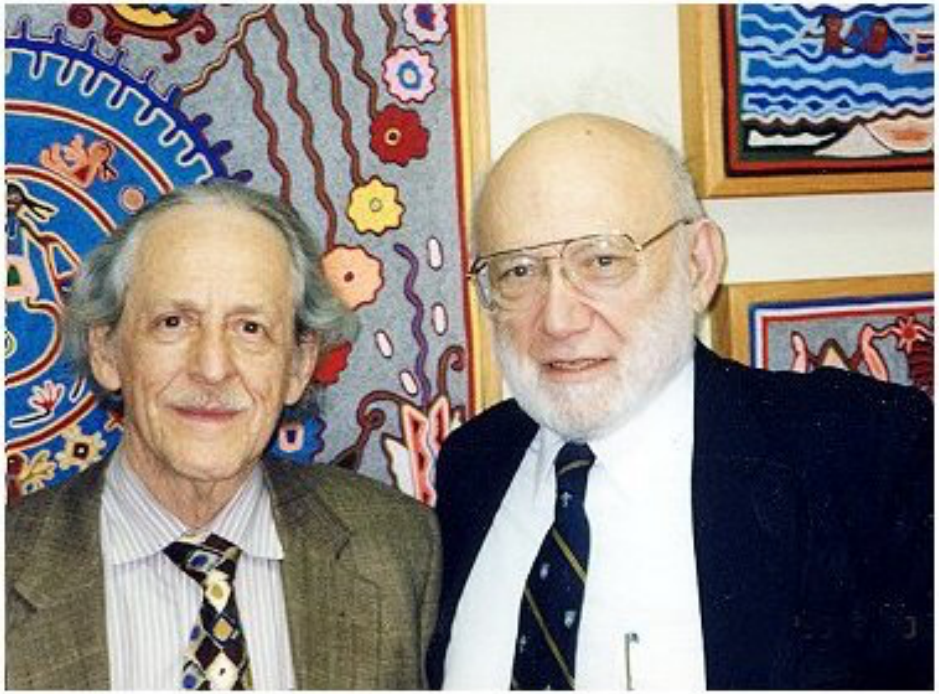
figshare
credit for all your research

ImpactStory.

MENDELEY

ResearchGate
Scientific Network

研究者が自身の専門知識と技能を使って「使いたいサービス」を事業化



Eugene Garfield and Joshua Lederberg

1958年にノーベル賞を受賞したJoshua Lederbergは、翌年Garfieldへの私信の中で政府助成金の取得を促した。

Genetics Citation Index（1961年版）はNIHとNSFの助成を受けて2年かかって完成し、最初のサイテーションインデックスとなった。

http://www.garfield.library.upenn.edu/lederberg/050959.html

MEMO FROM
J. LEDERBERG
GENETICS DEPARTMENT
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

TO: Eugene Garfield

TX
RC
RO

5/9/59.

Since you first published your scheme for a "citation index" in Science about 4 years ago I have been thinking very seriously about it, and must admit I am completely sold. In the nature of my work I have to spend a fair amount of effort in reading the literature of collateral fields and it is infuriating how often I have been stumped in trying to update a topic, where your scheme would have been just the solution! I am sure you critics have simply not grasped the idea, & especially the point that the author must learn to cooperate by his own choice of citations + then he does the critical work.

Have you tried to get this out in an adequate experiment? Would you look for support from the NSF? Of course you have to count on opposition from the established outfits, which have already succeeded in blocking any progressive centralization of the Index Medicus.

Sincerely

Joshua Lederberg

（再び）問題意識

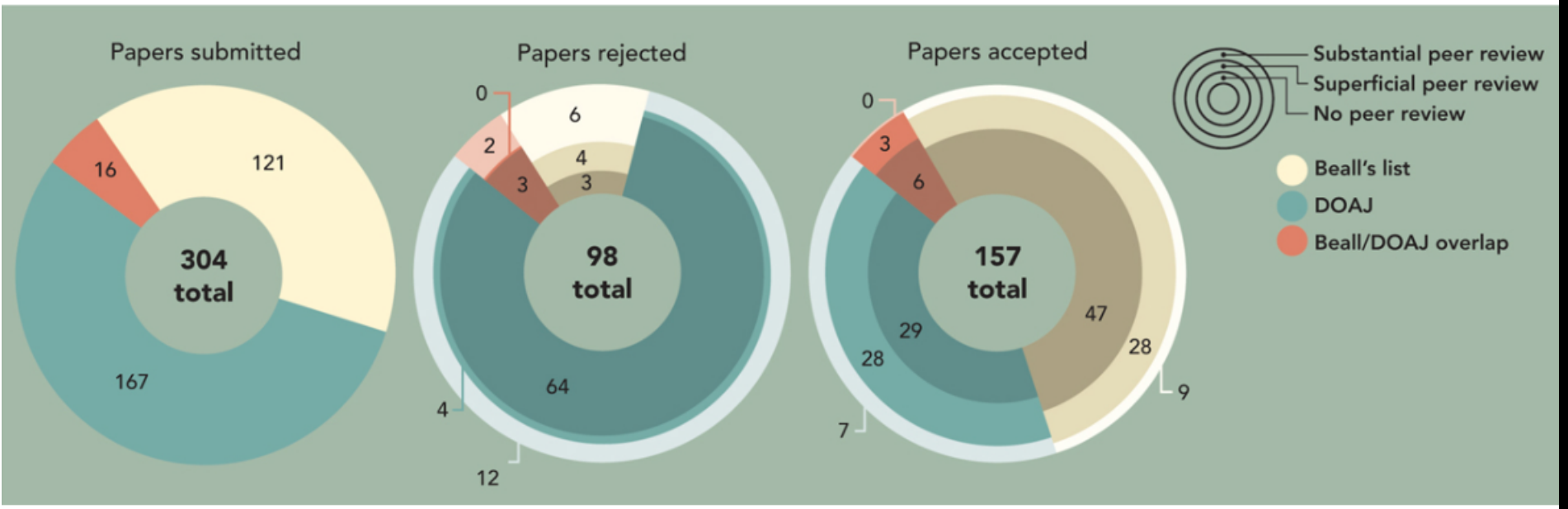
1. 「学術コミュニケーション」と「技術革新」
   - 情報技術の発達は、学術コミュニケーションの変容を促した。
     - 成果からプロセスの開示へ
     - よりオープンなアクセスへ
     - 正当性の審査から説明責任の確立へ
     - 受益者負担から発信者負担へ
     - 閉ざされたラボラトリーからよりソーシャルな共同作業へ
2. 「量」と「質」
   - 学術研究が国家予算によって支えられるようになり、大規模かつより多くの研究アウトプットを生んだ。＞競争と選択
   - オープンな学術コミュニケーションは、多くの新しいデータの取得を可能とし、「前倒し」かつ多様な評価が可能になった。
3. 学術コミュニケーションにおける「アクター」
   - 一般市民の科学プロジェクト参加や、研究者による学術情報ツールの開発によって、科学の担い手が多様化した。

人工知能はどのように学術コミュニケーションを変えていくのか？

あるいは変えつつあるのか？

Who's Afraid of Peer Review?

A spoof paper concocted by *Science* reveals little or no scrutiny at many open-access journals



Bohannon, J. Who's Afraid of Peer Review? *Science (New York, N.Y.)* **342**, 60–65 (2013).

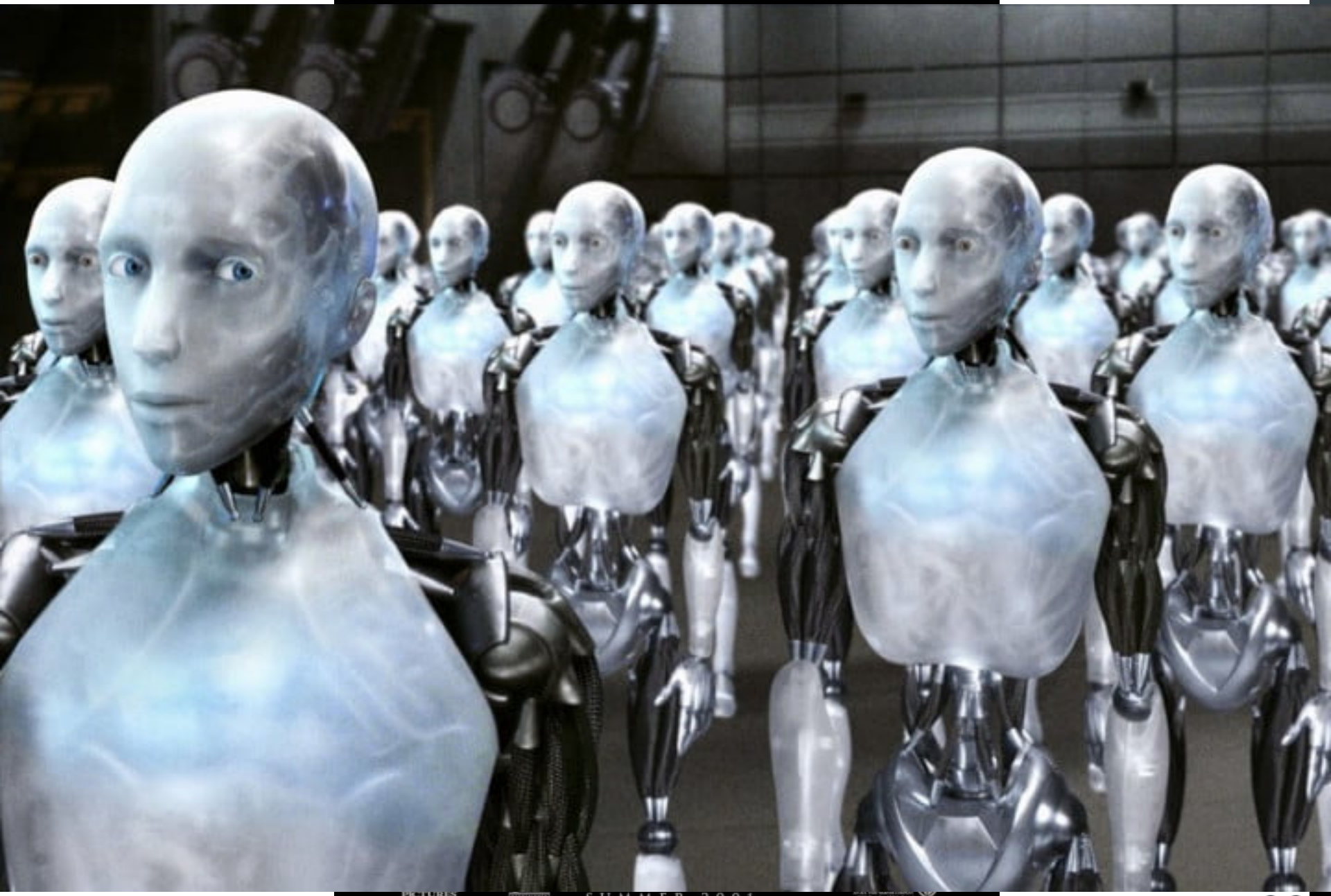
artificial intelligence noun

Definition of *artificial intelligence*

1 : a branch of computer science dealing with the simulation of intelligent behavior in computers

2 : the capability of a machine to imitate intelligent human behavior

https://www.merriam-webster.com/dictionary/artificial%20intelligence


PICTURES
SUMMER 2001
America Online Keyword: A.I. www.Aimovie.com

AI in scholarly communication

もしAIが「高度に知的な人間の行動を模倣する（imitate intelligent human behavior）」ものだとしたら、学術コミュニケーションの世界では一体何が起こるのか？

What do you mean by AI?

AI and machine learning are often used interchangeably, especially in the realm of big data. But these aren't the same thing, and it is important to understand how these can be applied differently.

- Artificial intelligence is a broader concept than machine learning, which addresses the use of computers to mimic the cognitive functions of humans.
- Machine learning is a subset of AI and focuses on the ability of machines to receive a set of data and learn for themselves, changing algorithms as they learn more about the information they are processing.
- Deep learning goes yet another level deeper and can be considered a subset of machine learning. The concept of deep learning is sometimes just referred to as "deep neural networks," referring to the many layers involved. A neural network may only have a single layer of data, while a deep neural network has two or more.

https://www.datasciencecentral.com/profiles/blogs/artificial-intelligence-vs-machine-learning-vs-deep-learning

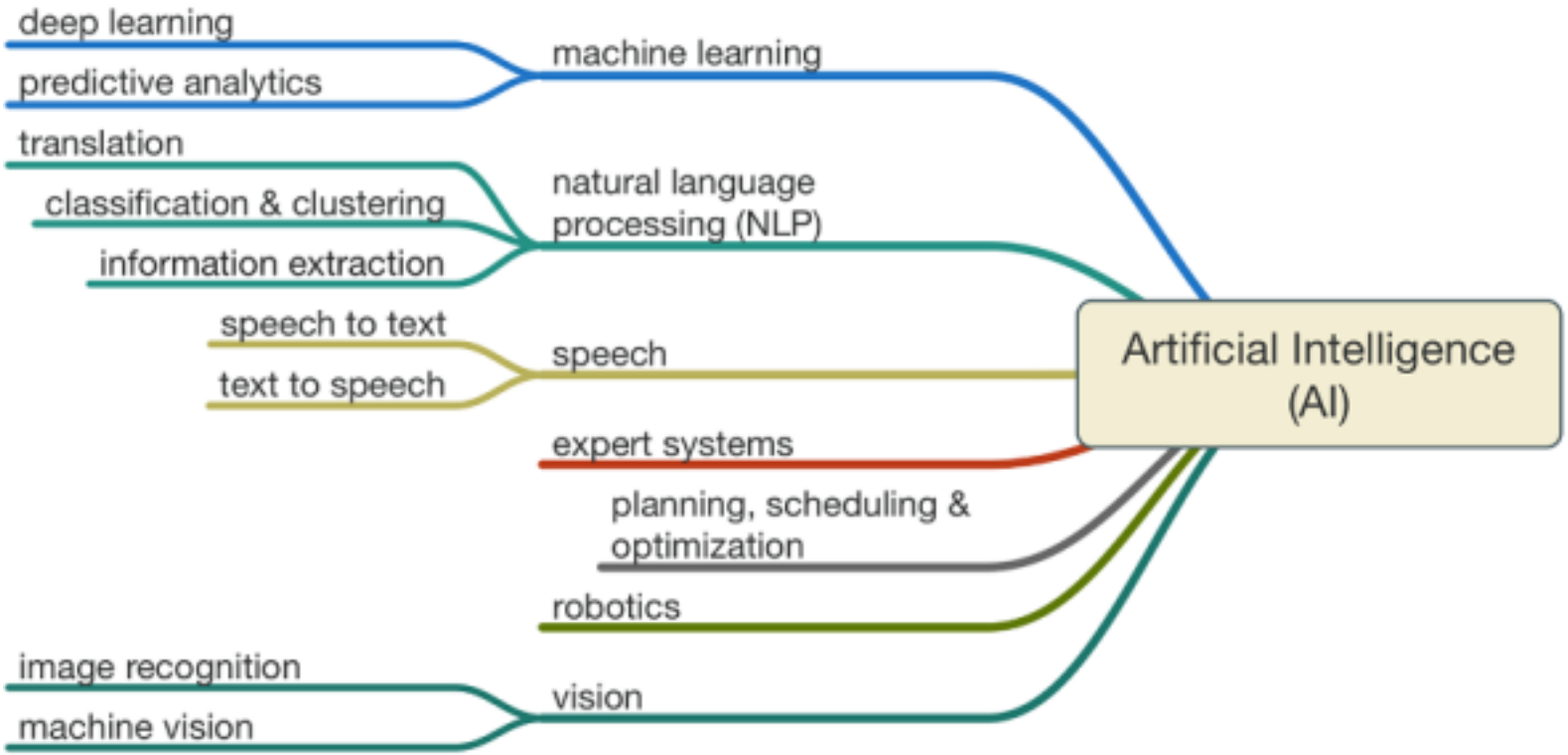
What do you mean by AI?

AI（人工知能）と 機械学習は、特にビッグデータの分野ではしばしば互換可能な用語として使われています。しかし、これらは同じものではなく、どのように異なるのかを理解することは大切です。

- 人工知能（Artificial intelligence）は機械学習よりもより広義な用語で、人間の認知機能を模倣するためにコンピュータを利用することを指します。
- 機械学習（Machine learning）は人工知能の下層概念で、機械が一定のデータを受理して、情報を処理する過程で学習しながらアルゴリズムを変更することができるという能力を指します。
- 深層学習（Deep learning）はさらに下層の概念で、機械学習の一部です。深層学習のコンセプトは、単に多層のニューラルネットワークを利用しているということを指すこともあります。ニューラルネットワークは単層のデータでも成立しますが、ディープニューラルネットワークは2層以上のデータから成ります。

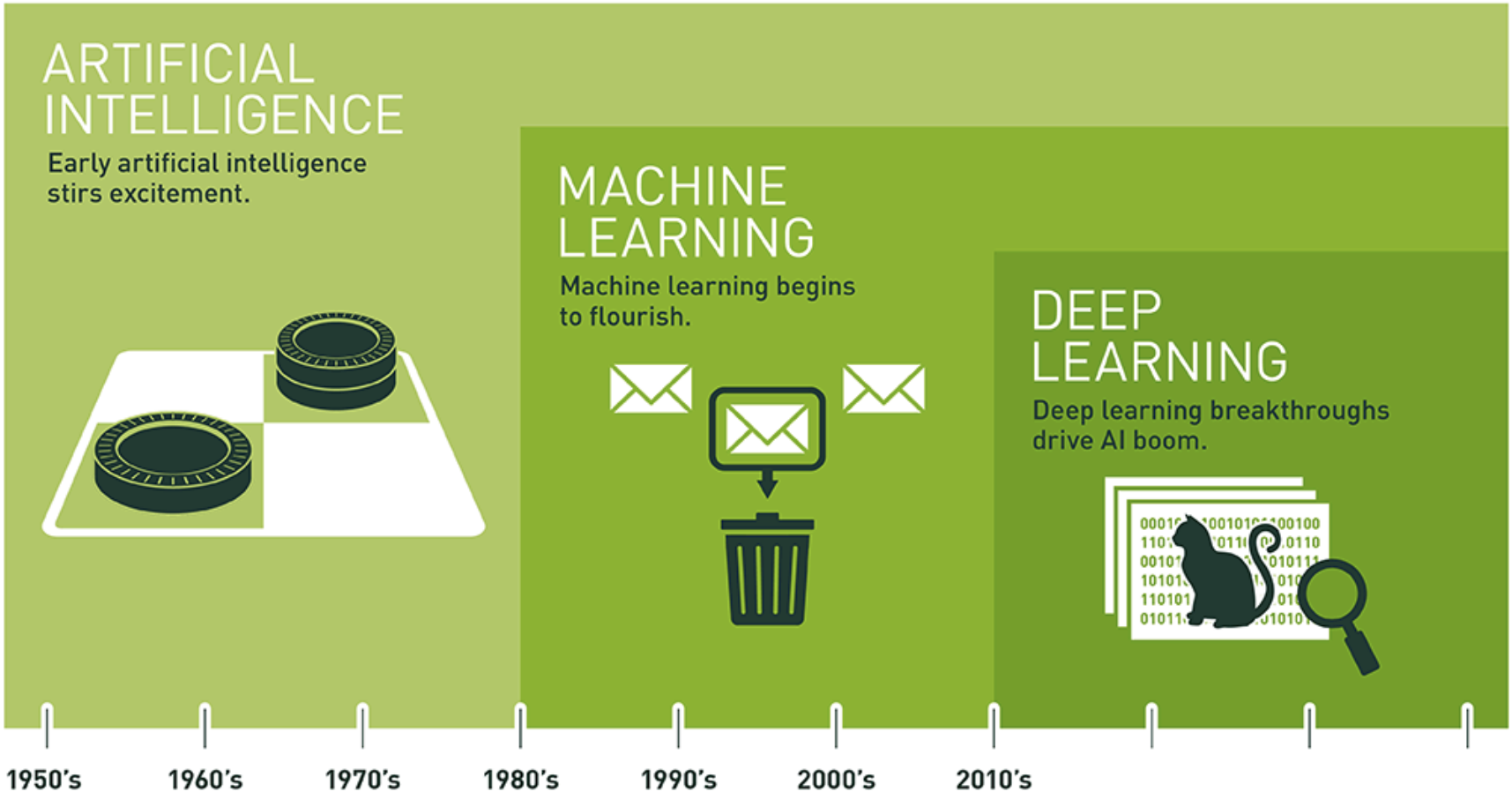
https://www.datasciencecentral.com/profiles/blogs/artificial-intelligence-vs-machine-learning-vs-deep-learning

Different branches of AI



https://ilovemyarchitect.com/tag/machine-learning/

They have been around for a while…



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

https://www.datasciencecentral.com/profiles/blogs/artificial-intelligence-vs-machine-learning-vs-deep-learning

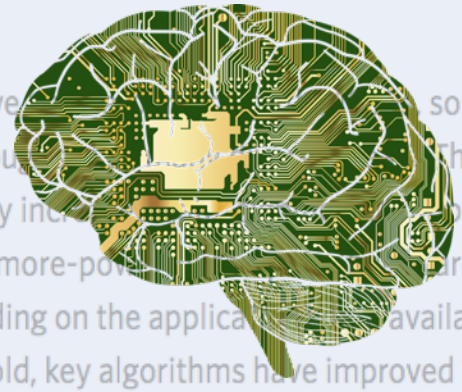
Why now?

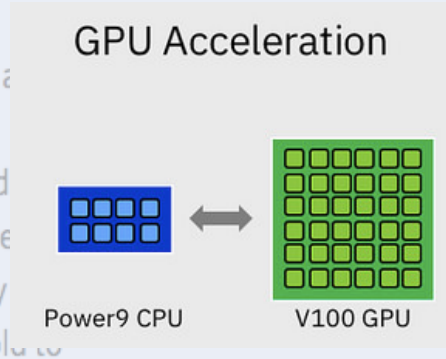
Harvard Business Review | THE BIG IDEA

WHAT'S DRIVING THE MACHINE LEARNING EXPLOSION?

BY ERIK BRYNJOLFSSON AND ANDREW MCAFEE

JULY 18, 2017





https://hbr.org/2017/07/whats-driving-the-machine-learning-explosion

CLOUD MACHINE LEARNING SERVICES COMPARISON

	Amazon	Microsoft	Google	IBM
Automated and semi-automated ML services				
	Amazon ML	Microsoft Azure ML Studio	Google Prediction API	IBM Watson ML Model Builder
Classification	✓	✓	deprecated	✓
Regression	✓	✓		✓
Clustering	✓	✓		✗
Anomaly detection	✗	✓		✗
Recommendation	✗	✓		✗
Ranking	✗	✓		✗
Platforms for custom modeling				
	Amazon SageMaker	Azure ML Services	Google ML Engine	IBM Watson ML Studio
Built-in algorithms	✓	✗	✗	✓
Supported frameworks	TensorFlow, MXNet, Keras, Gluon. Pytorch, Caffe2, Chainer, Torch	TensorFlow, scikit-learn, Microsoft Cognitive Toolkit, Spark ML	TensorFlow, scikit-learn, XGBoost, Keras	TensorFlow, Spark MLlib, scikit-learn, XGBoost, PyTorch, IBM SPSS, PMML

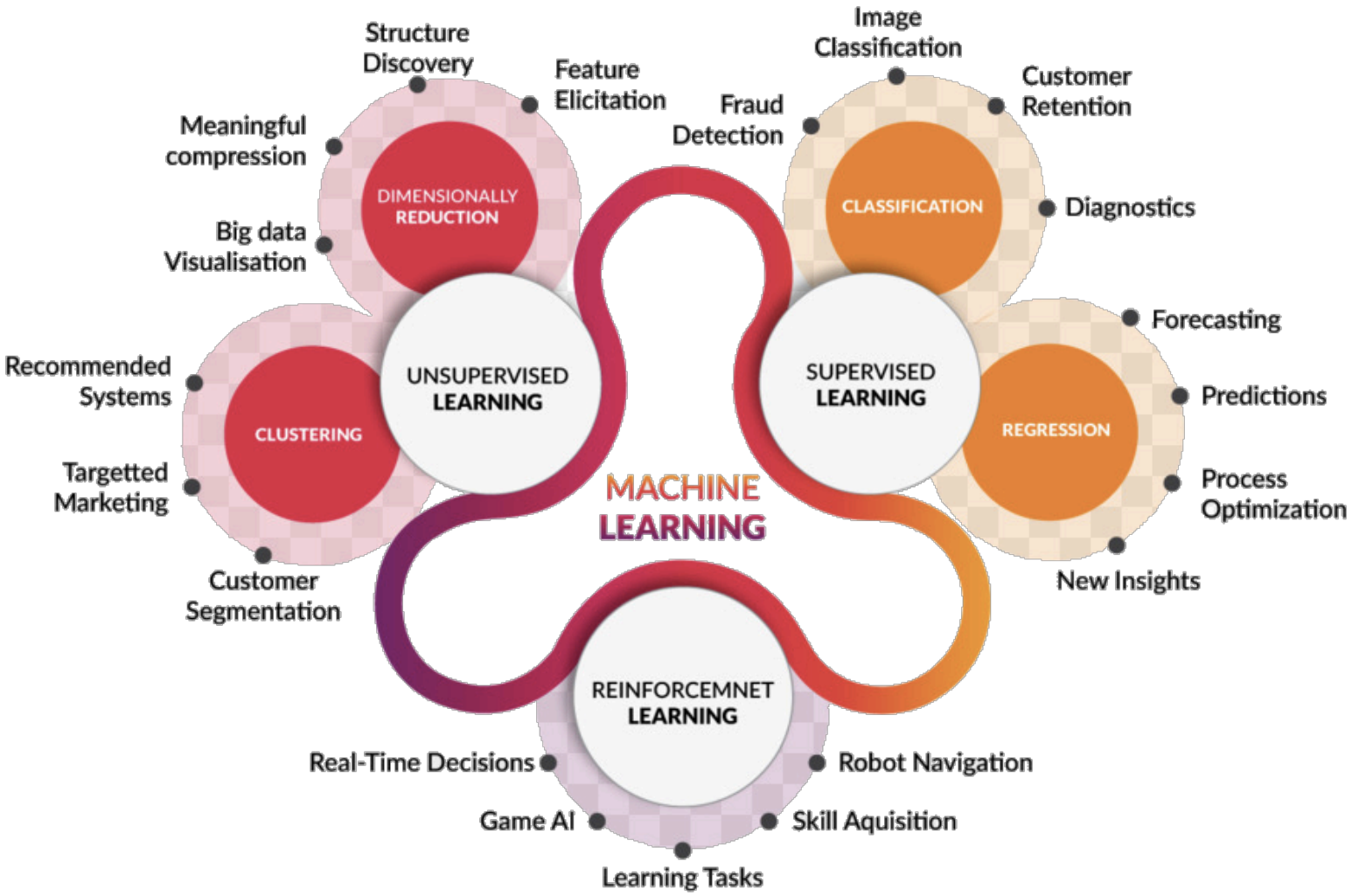
MLaas?

Machine learning as a service (MLaaS) is an array of services that provide machine learning tools as part of cloud computing services. MLaaS helps clients benefit from machine learning without the cognate cost, time and risk of establishing an inhouse internal machine learning team.

altexsoft
software r&d engineering

https://www.altexsoft.com/blog/datascience/comparing-machine-learning-as-a-service-amazon-microsoft-azure-google-cloud-ai-ibm-watson/

Machine Learning applications



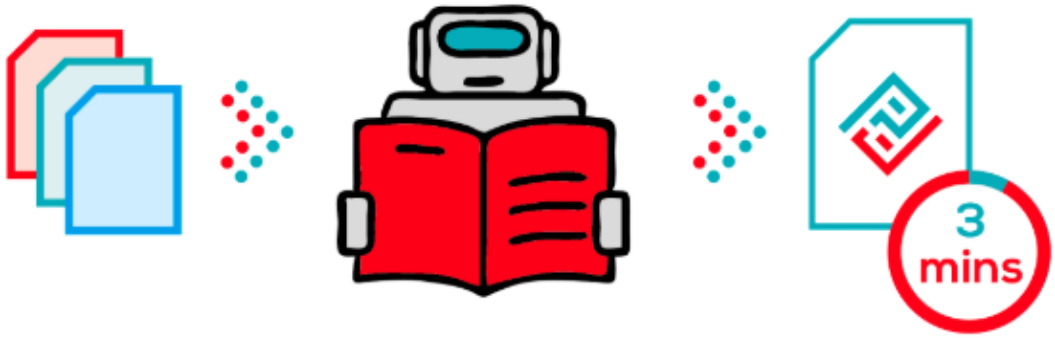
http://www.researchinfinitesolutions.com/blog/Machine-learning-Algorithms

Case study: paper digest

学術論文の「ダイジェスト」（抽出型要約）を自動生成するサービス

Artificial Intelligence summarizes
academic articles for you

Reducing reading time to ***3 minutes!***

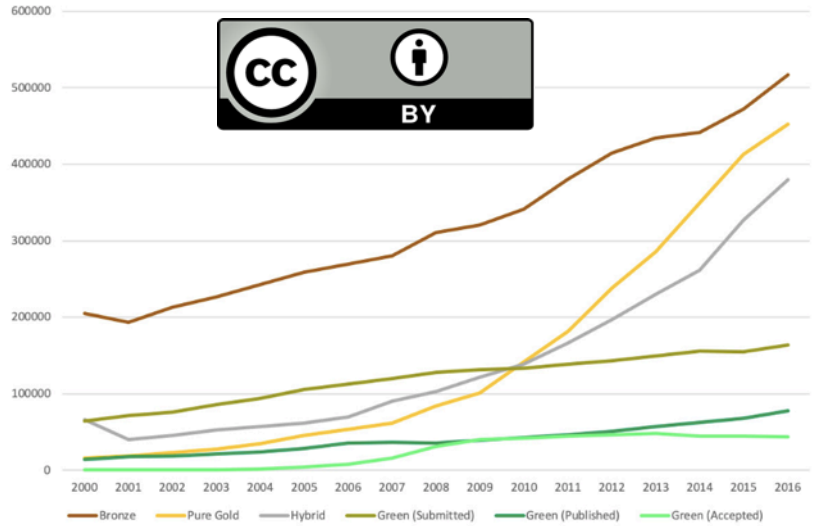


https://www.paper-digest.com

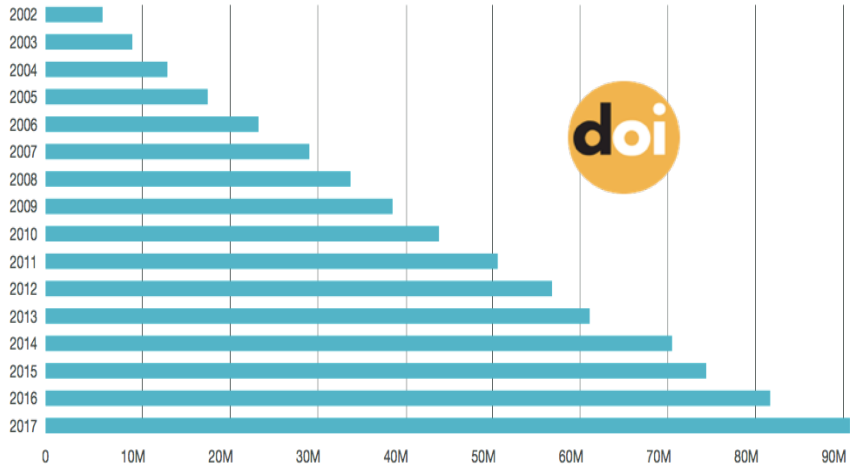
Paper Digestとは

- Paper Digestは、二人のポスドクが自身の経験に基づいて、自らのスキルを使って課題の克服に取り組んだ実験的プロジェクトです。
- 偶然彼らと出会った私は、20年近くSTM業界で働いた経験を通じて、このプロジェクトを私的に支援しています。
- Digital Science社がスタートアップを支援するCatalyst Grantを受賞し、事業化への道を模索しています。

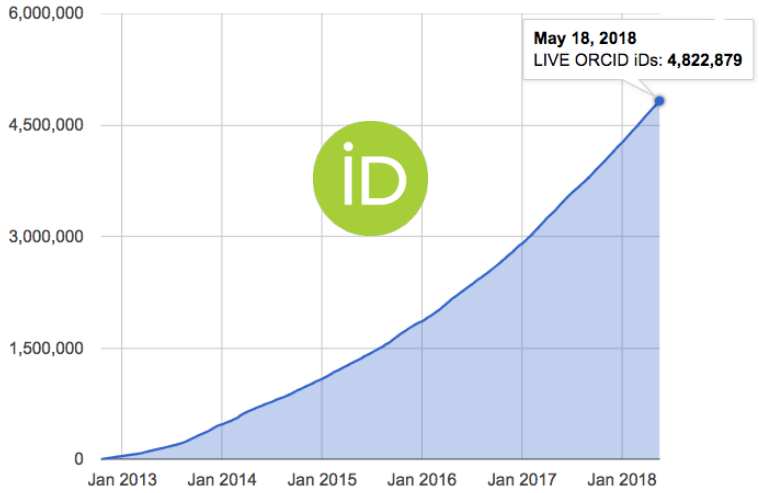
paper digest



The Ascent of Open Access
https://doi.org/10.6084/m9.figshare.7618751.v2



Crossref. 2016-17 annual report.
https://www.crossref.org/pdfs/annual-report-2016.pdf



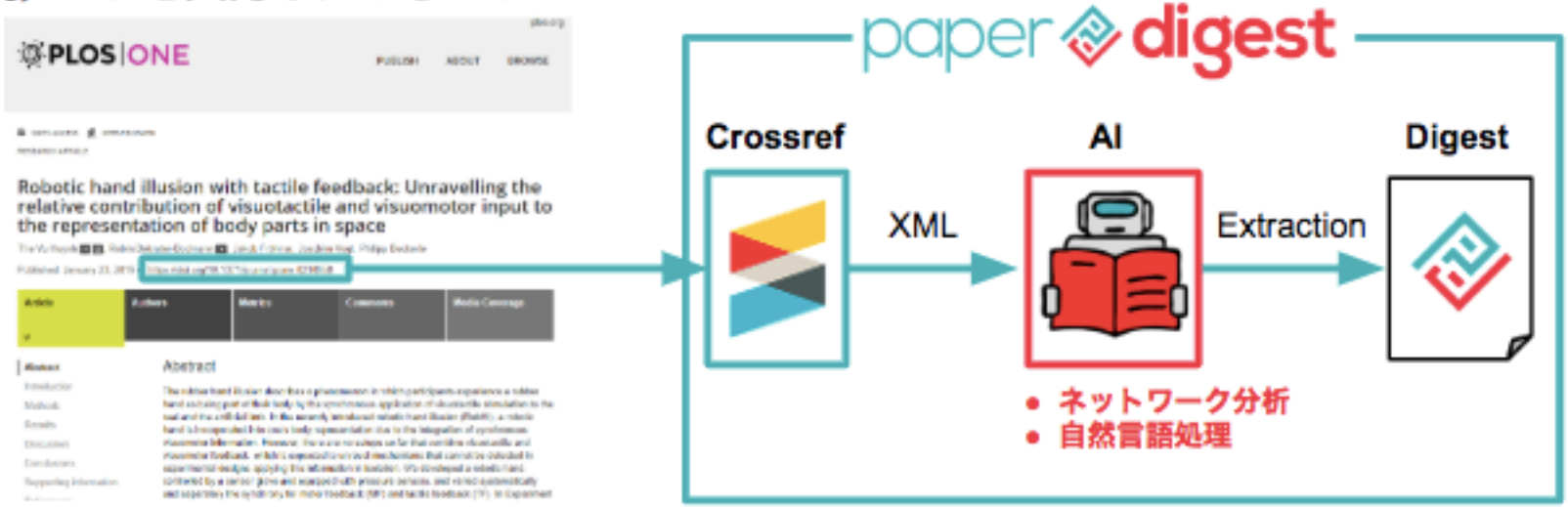
ORCID live iDs
https://support.orcid.org/knowledgebase/articles/150557-number-of-orcid-ids

人工知能が処理できる「機械可読なメタデータ」が付いている学術情報はどんどん増えている。

実際の処理フロー

OAかどうかを確認→本文を取得→各文の**重要度**を計算してスコア化
→スコアが高いものを抽出（＝Digest）

例）DOIを入力したときのフロー



オープンなコンテンツは、人間だけでなく機械にとっても無料で読めるもの。
ただし、「機械可読なメタデータ」が必要。

Paper Digest wins SSP Previews Session People's Choice Award

June 12, 2019 – Mount Laurel, NJ – At the Society for Scholarly Publishing (SSP) 41st Annual Meeting held on May 29-31, in San Diego, CA, Paper Digest was awarded a People's Choice Award for their artificial intelligence (AI) tool for scholarly communications during the *SSP Previews Session, New and Noteworthy Product Presentations.*

Pictured from left, David Myers, Dr. Cristian Mejia, Nobuko Miyairi, Dr. Yasutomo Takano, and SSP President, Adrian Stanley

This "lightning-round" plenary offered the chance to learn more about the industry's newest and most innovative products, platforms, and/or content in 5-minute, back-to-back presentations. Attendees left with exposure to a wide breadth of content on new and updated products of interest to the scholarly publishing community. After the 13 presentations, attendees were asked to vote via the mobile meeting app for the product, platform, service, or content that would likely have the most positive impact in scholarly communications.

Paper DigestはSSP第41回年次大会の新製品部門で得票数1位となりました。

https://www.sspnet.org/community/news/paper-digest-wins-ssp-previews-session-peoples-choice-award/

AI in scholarly communication

もしAIが「高度に知的な人間の行動を模倣する（imitate intelligent human behavior）」ものだとしたら、学術コミュニケーションの世界では一体何が起こるのか？

Paper Digestの場合：

1. 読みたい論文のライセンスを判断する
2. OAなら本文にアクセスできる
3. 自動で要約する
    1) 必要なら全部読む
    2) 不要なら読まない

AI in scholarly communication

もしAIが「高度に知的な人間の行動を模倣する（imitate intelligent human behavior）」ものだとしたら、学術コミュニケーションの世界では一体何が起こるのか？

「検索」から「発見」へ

「全部読む」から「まとめて理解する」へ

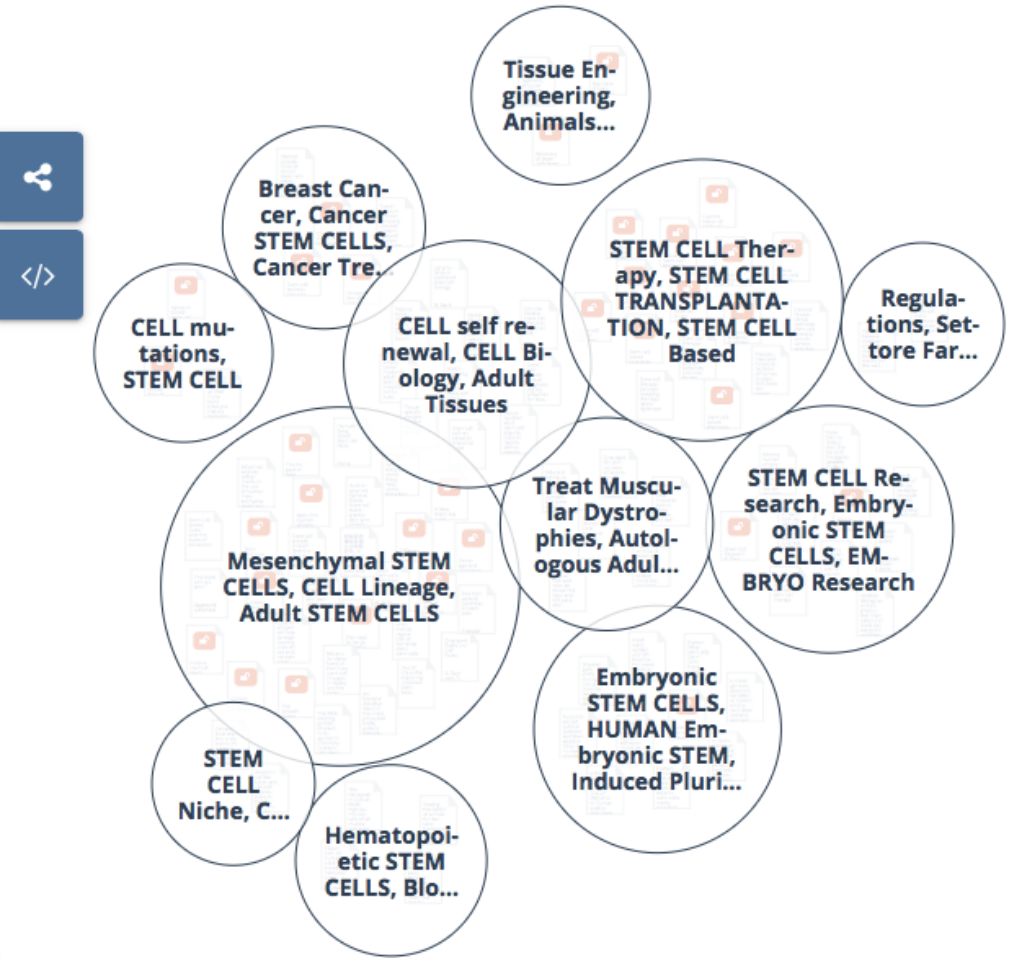
「自分で書く」から「機械が書く」へ

ディスカバリー系

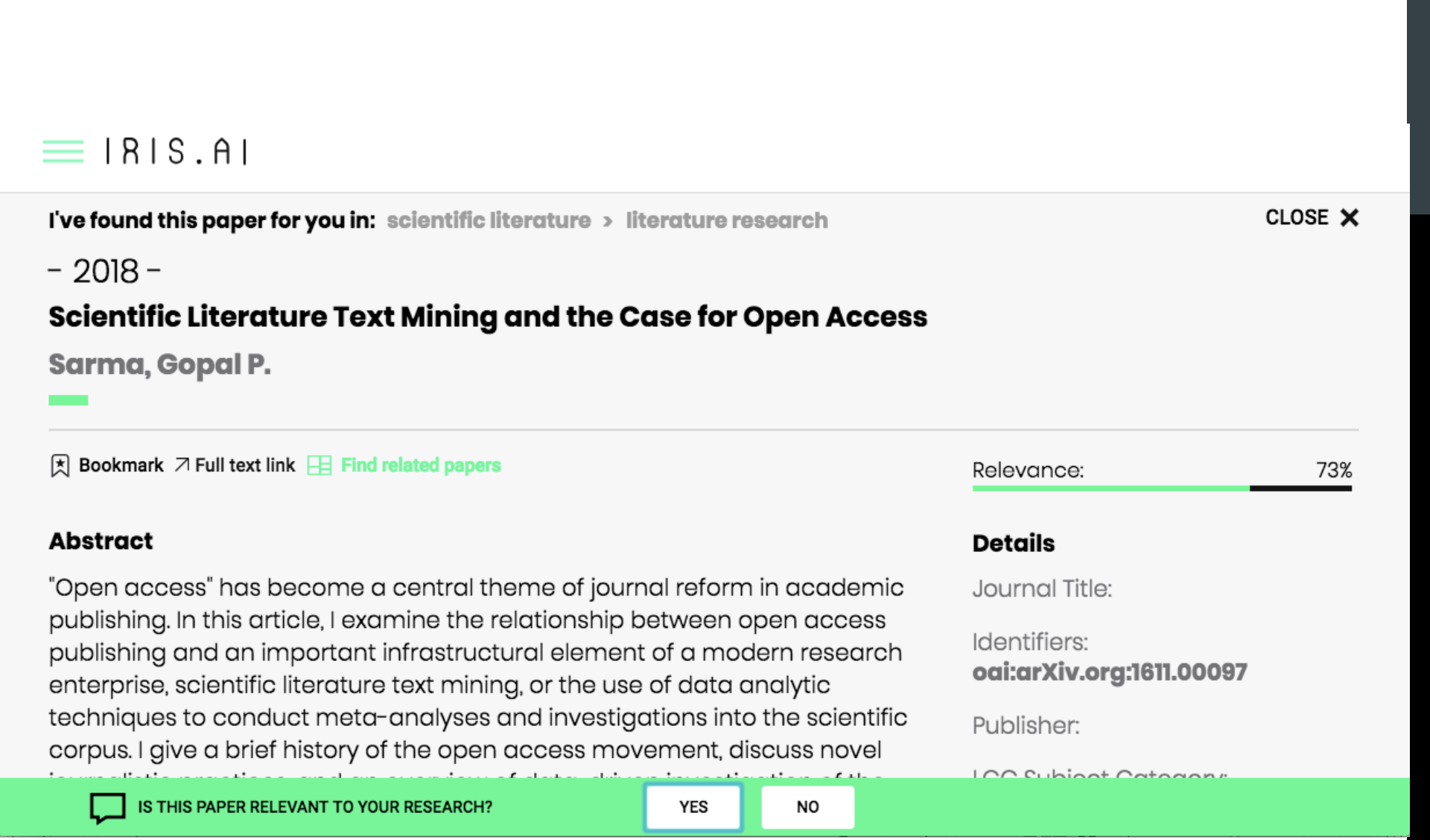
オリジナルのコンテンツ（論文、データなど）や研究情報を「発見」するもの。人間が「検索」するためにはキーワードを考える必要があるが、機械は自然言語処理やテキストクラスタリング、協調フィルタリングなどを使って「似たようなもの」を探すのが得意。人間が見つけにくいものを見つける。

例：

- Open Knowledge Map https://openknowledgemaps.org/
- Iris.ai https://iris.ai/
- Dimensions https://www.dimensions.ai/
- Sparrho https://www.sparrho.com/
- Quatolio https://quartolio.com/
- Get the research https://gettheresearch.org/



「似ている」論文をクラスタリングしてグラフィック表示



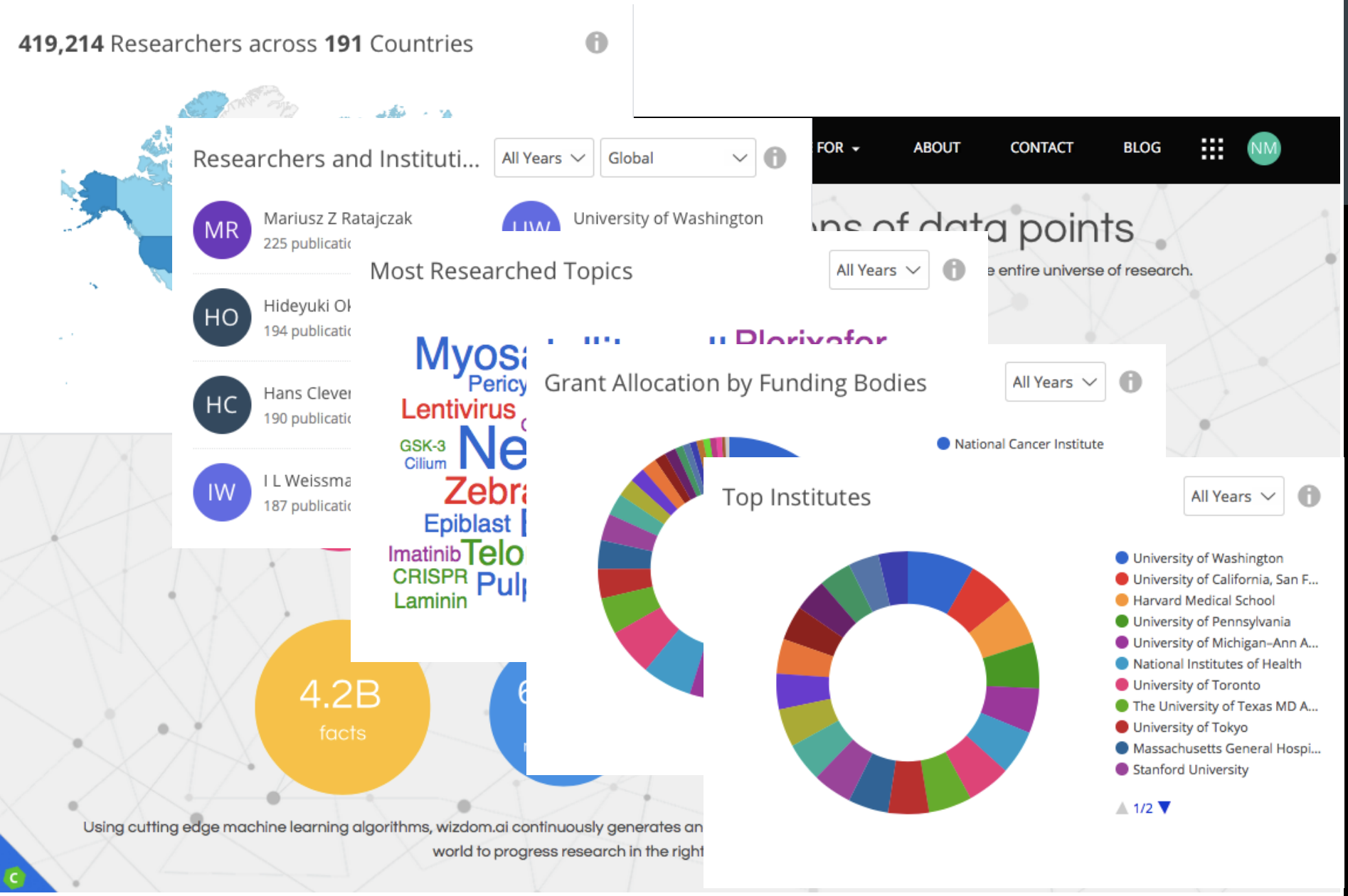
シード（種）となる論文やプレゼンテーションを入力すると、それに似たものをグラフィック表示> クリックして絞り込み

要約・まとめ系

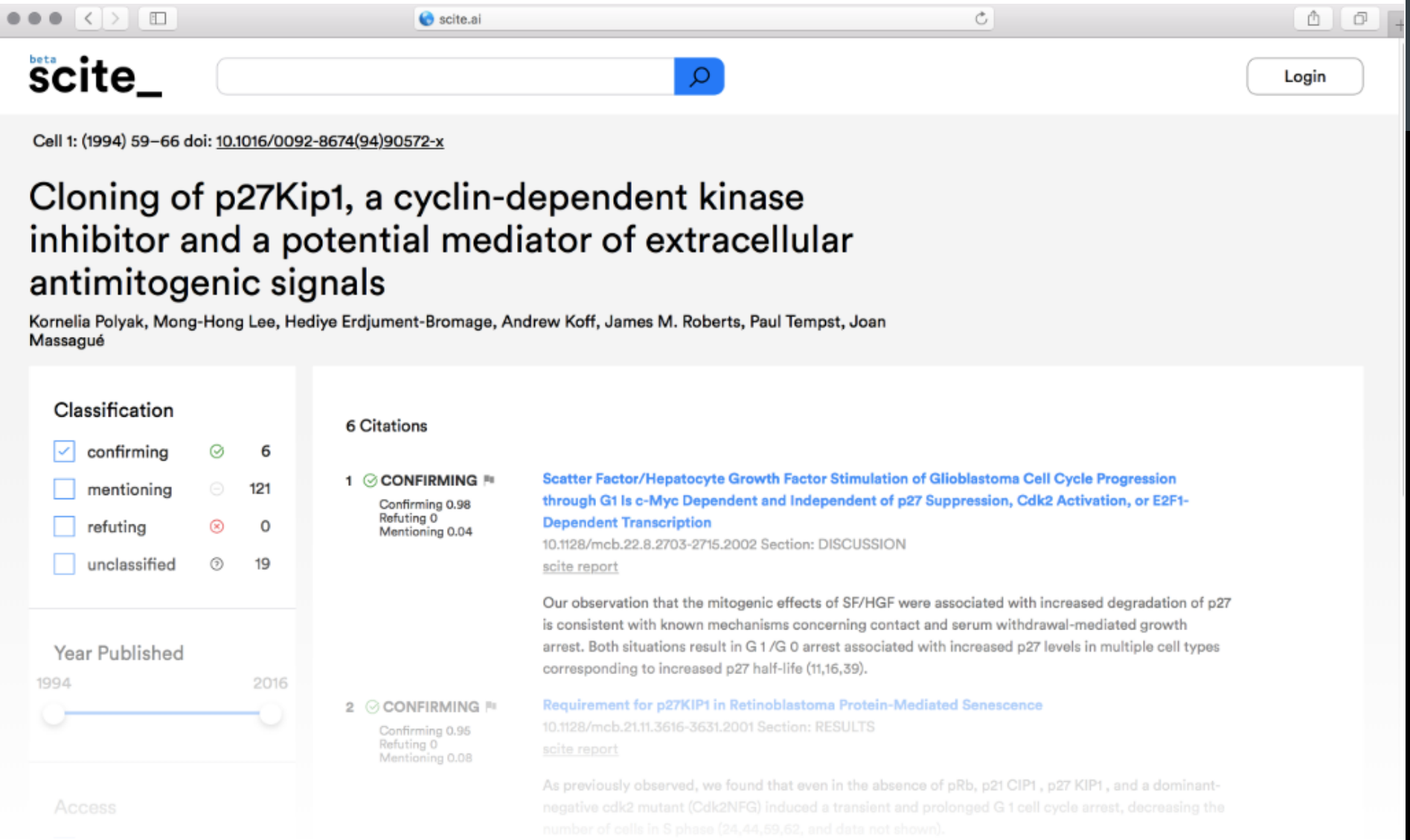
オリジナルコンテンツを要約したり、追加情報により文脈を与えたりするもの。テキスト抽出や類型化により、より知的な情報を付加する。情報が多すぎるときに便利。

例：

- Paper Digest https://www.paper-digest.com
- Scholarcy https://www.scholarcy.com/
- Wizdom.ai https://www.wizdom.ai/
- Sci.ai https://sci.ai/



キーワード検索の結果を、グラフィカルに要約したフォーマットで表示。＞クリックして絞り込み



既知の論文がどのように引用されているかを自動判定。＞肯定的、否定的、単なる言及、など

書いてくれる系

機械が書いてくれる、もしくは人間の執筆作業の工程（文献レビュー、原稿校正、フォーマッティングなど）を助けてくれるもの。時間がかかるマニュアル作業を減らしてくれる。

例：

- Penelope.ai https://www.penelope.ai/
- RAx https://raxter.io/
- AI Writer http://www.ai-writer.com

You have a results section.

Where no a priori power calculations have been provided, inferential statistics (e.g. p-values) are not appropriate and should not be presented; however, descriptive statistics (e.g. averages, variances/confidence intervals) are appropriate and encouraged.

This section should be titled Conclusions.

You have a conclusions section. Remember to focus on the primary conclusions and implications. Do not go beyond the data.

Additional Headings

You have included an acknowledgements section.

You have included a section about funding.

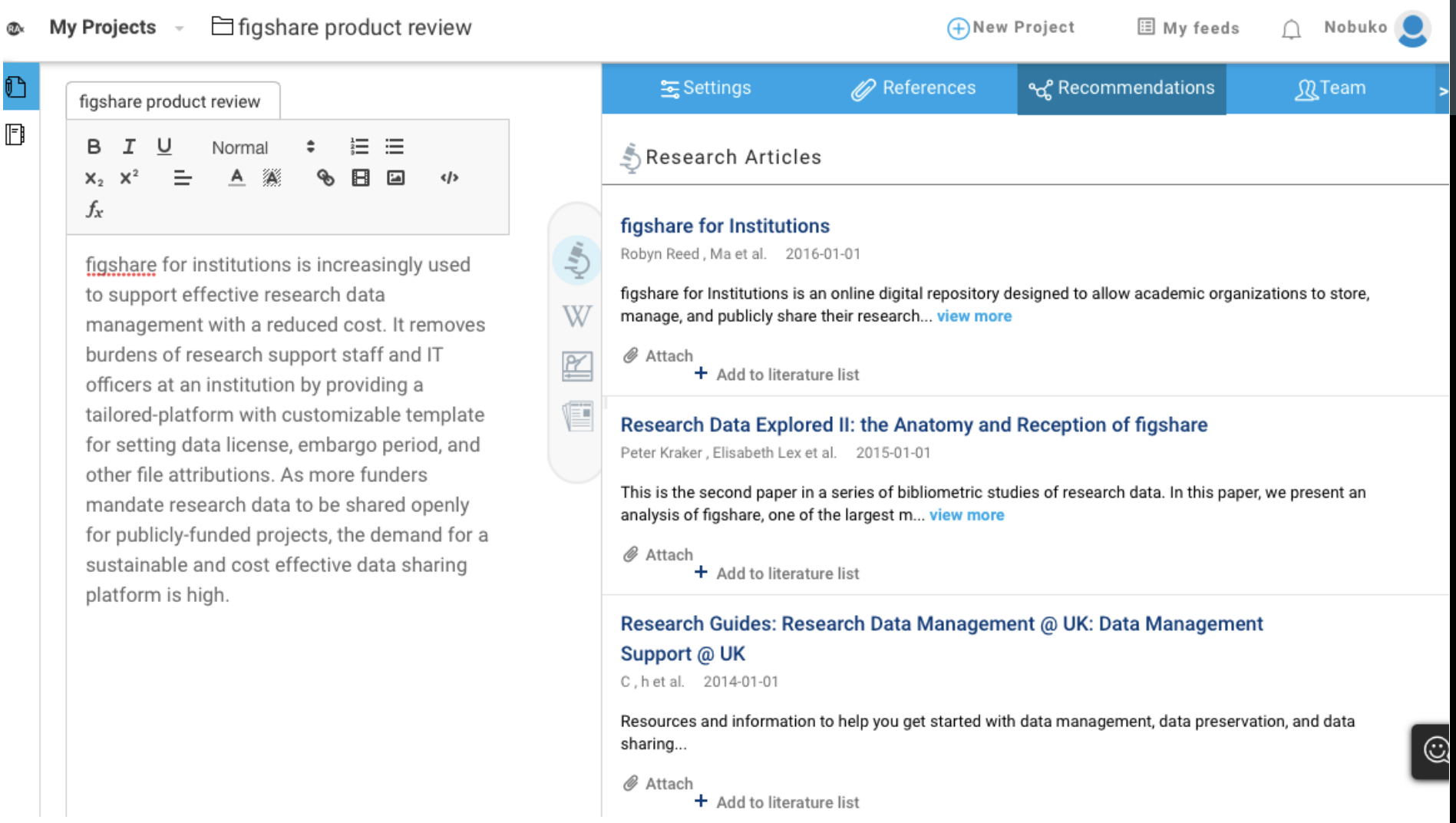
BMJ Open Science suggests the following wording: either: 'This work was supported by [name of funder] grant number [xxx]' or 'This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors'. You must ensure that the full, correct details of your funder(s) and any relevant grant numbers are included.

diffusion of these compounds throughout the recorded MSN. The presence of MK-801 (Figure 5C) or EGTA (Figure 5D) blocked the induction of long-term depression at Pf synapses. These results thus suggest that Pf synapses exhibit long term plasticity that is postsynaptically mediated by calcium flowing through NMDA receptors.

Discussion

The main findings of the present study are that thalamostriatal synapses with MSNs originating in different parts of the intralaminar thalamus have markedly different properties. Using an optogenetic approach in mice we were able to dissociate between the rostral (CL) and caudal (Pf) intralaminar inputs to MSNs. First, we find that CL synapses give rise to larger amplitude responses than Pf synapses. Secondly, CL synapses are facilitating, whereas Pf synapses are depressing. Thirdly, CL synapses predominantly express postsynaptic AMPA receptors, whereas Pf synapses are predominantly associated with postsynaptic NMDA receptors. Finally, the re

原稿をアップロードすると、どこを直したらいいか分析してチェックリストを返してくれる。



原稿を書いている最中に、その内容から過去の関連文献を自動検索して引用すべき文献としてリコメンドしてくれる。

Back home

Restart tutorial - Back to AI-Writer.com

Article | Raw Sentences | Sources

Rewrite circa 53% of the original text.

Highlight text that has not been rewritten

How Does Machine Learning Work

Data mining uses a lot of machine learning methods, but with different objectives : on the other hand, machine learning also uses the method of" unattended education" or as a pre-programmed step to improve the accuracy of the learner .[11]

Generalisation in the context is the ability of a learning machine to perform precisely on new and invisible examples after the experience of a data set .[11]

Although machine learning has been transformed into some fields, it is difficult to learn effectively because the search for patterns is difficult and often not enough training data are available, as a result, many machine learning programmes are not able to provide the desired value .[11]

Due to such challenges, the effective use of automatic learning can take longer to be used in other areas .[11]

Amazon wise allows data scientists and developers to quickly and easily build, train and deploy machine learning models with high-performance automated machine learning algorithms, wide-frame support and a-click workout, tune-up and conclusions .[33]

New machine learning developers will find the interface more familiar with the traditional code, as machine learning templates can be defined and manipulated like any other data structure .[33]

Machine learning requires a wide range of powerful computing options, from GPUs to computer-intensive learning, to FPGAs for specialised hardware acceleration, to high-memory instances for ongoing inference .[33]

Cited Sources

https://martechtoday.com/how-machine-learning-works-150366 [0]

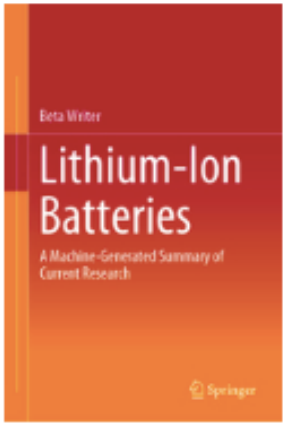
https://machinelearningmastery.com/how-machine-learning-algorithms-work/ [1]

引用しようと思っている文献と、書こうとしているトピックを入力すると、AIが文章を書いてくれる。

Springer Nature publishes its first machine-generated book

Innovative book prototype provides a compelling machine-generated overview about the latest research on lithium-ion batteries, automatically compiled by an algorithm developed in collaboration with the Applied Computational Linguistics lab of Goethe University Frankfurt/Main (Germany)

London | Heidelberg, 02 April 2019



Springer Nature published its first machine-generated book in chemistry. The book prototype provides an overview of the latest research in the rapidly growing field of lithium-ion batteries. The content is a cross-corpus auto-summarization of a large number of current research articles in this discipline. Serving as a structured excerpt from a huge set of papers, the innovative pipeline architecture aims at helping researchers to manage the information overload in this discipline efficiently.

In close collaboration between Springer Nature and researchers from Goethe University Frankfurt/Main, a state-of-the-art algorithm, the so-called *Beta Writer*, was developed to select, consume and process relevant publications in this field from Springer Nature's content platform *SpringerLink*. Based on this peer-reviewed and published content, the Beta Writer uses a similarity-based clustering routine to arrange the source documents into coherent chapters and sections. It then creates succinct summaries of the articles. The extracted quotes are referenced by hyperlinks which allow readers to further explore the original source documents. Automatically created introductions, table of contents and references facilitate the orientation within the book.

今年の4月、初の「人工知能が書いた本」が出版された。

https://group.springernature.com/la/group/media/press-releases/springer-nature-machine-generated-book/16590134

New AI fake text generator may be too dangerous to release, say creators

The Elon Musk-backed nonprofit company OpenAI declines to release research publicly for fear of misuse

▲ The AI wrote a new passage of fiction set in China after being fed the opening line of Nineteen Eighty-Four by George Orwell (pictured). Photograph: Mondadori/Getty Images

あまりにももっともらしい文章を書く人工知能は「危険」なので非公開に。

https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction

Your next press release will not be written by a robot [Beyond the Journal: The science of communication]

Fears that automation will take our jobs are rife, but press officers need not feel threatened by new tools that translate technical research into lay summaries.

A robot did not write this article.
Copyright : ktsdesign/123rf

By Ruth Francis

I recently attended a seminar that looked at tools we can use to help promote research. One of the panelists, Sabine Louët of Science Pod, sparked a discussion I had afterwards with two other press officers about the value we bring compared to what can be automated.

Louët showed examples of news content that was written by Natural Language Processing, a sub-discipline of artificial intelligence. For example, in 2014, the first news article reporting a small earthquake in Los Angeles, California was generated in this way, and others were created from crawling data on street crime or climatic events. For those of us who struggle with headline writing, A*STAR and Singapore Press Holdings in Singapore may be able to help: they are using AI to make headlines more engaging without resorting to pure clickbait.

At Science Pod, they have software that can augment the work of a press officer by automatically creating a lay summary of a research article. In this way, it demystifies the original research much like a press release.

プレスリリースも人工知能が自動で生成。でもやはり人間のプロにはかなわない？

https://www.researchsea.com/html/article.php/aid/12577/cid/6/research/people/researchsea/your_next_press_release_will_not_be_written_by_a_robot__beyond_the_journal__the_science_of_communication_.html

AI in scholarly communication

もしAIが「高度に知的な人間の行動を模倣する（imitate intelligent human behavior）」ものだとしたら、学術コミュニケーションの世界では一体何が起こるのか？

今のところ、新しく提供され始めているツールは以下のような目的に特化している。

- 探しているのに見つからないものを見つける（ディスカバリー系）
- 見つかりすぎて困るものをまとめてくれる（要約・まとめ系）
- 調べたり書いたりする手間を軽減してくれる（書いてくれる系）

機械が節約してくれる時間と労働力を、私たちは新たにどう使うのか？

情報プロフェッショナルの仕事はどのように変わっていくのか？

（最後に再び）問題意識

1. 「学術コミュニケーション」と「技術革新」
   - インターネットの普及と、それにともなうソーシャル化は、学術コミュニケーションをよりオープンに、透明にした。
   - AI（もしくは機械学習、深層学習）の学術情報ツールへの応用は、機械が得意なところから。最後まで残るものは？
2. 「量」と「質」
   - 「量」が増えたことは確実にAIの利用を促進したが、機械可読識別子の普及や、AI利用のコストダウンなどの背景事情によるところが大きい。
   - 機械に「質」の判断はできるのか？
3. 学術コミュニケーションにおける「アクター」
   - 生産者、利用者、仲介者
   - 機械はこのすべてになり得る可能性がある。

THANKS!

Nobuko Miyairi
nobuko.@miyairi.info