

# 学術コミュニケーションのエコシ ステムの今後 ～arXivの現状から考える～

武田英明  
国立情報学研究所  
takeda@nii.ac.jp

# arXivのMember Advisory Board (MAB) Meetingに行ってきました。



- 10/2 @Olin Library, Cornell University
- この会議はarXivのメンバーのうち、コンソーシアムなどの代表者などを招集して、1年に1度開催。
- 参加者
  - 多くが図書館：
    - Los Alamos Library, TIB, U of California, Ohio State Univ. U of Amsterdam, U of Queensland, U of Sydney
  - 研究助成機関
    - Jisc, Simons Foundation
  - 研究者代表: professor at KTH (欠席)

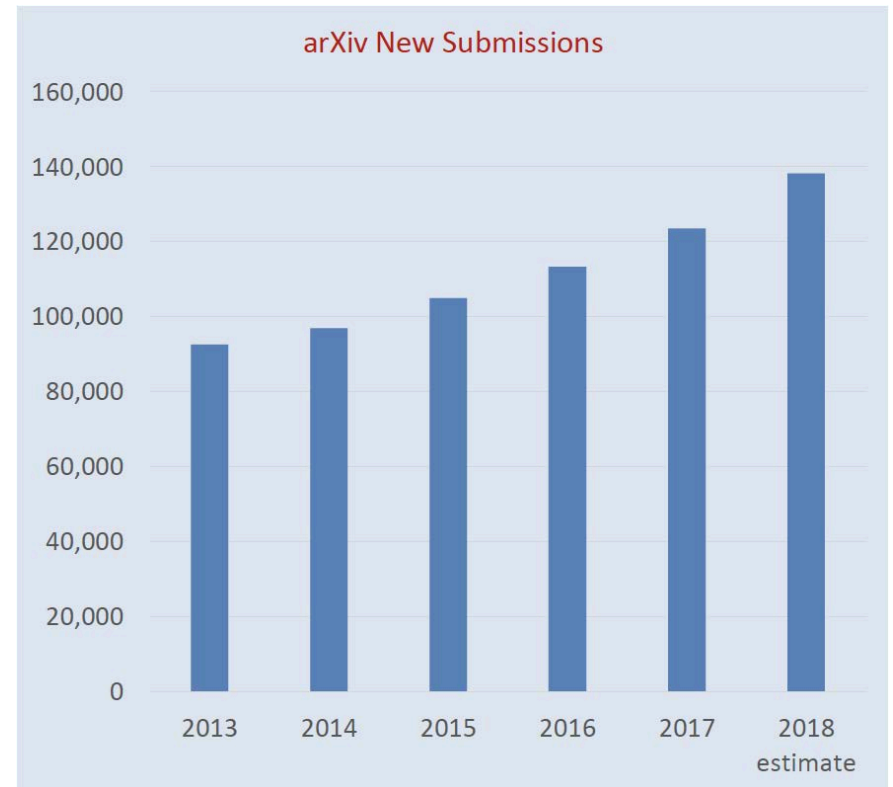


# arXivの現状

- 1,443,000 papers
- 600 new submissions/day
- 7 paper downloads/second
- 162 volunteer moderators
- 1 billion download+

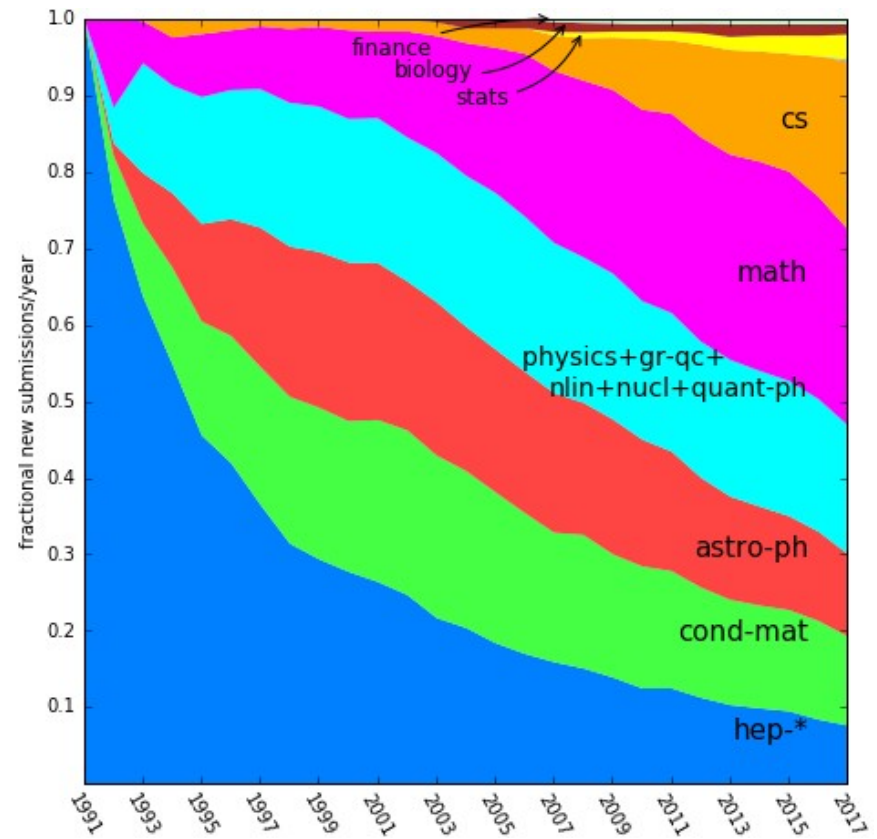
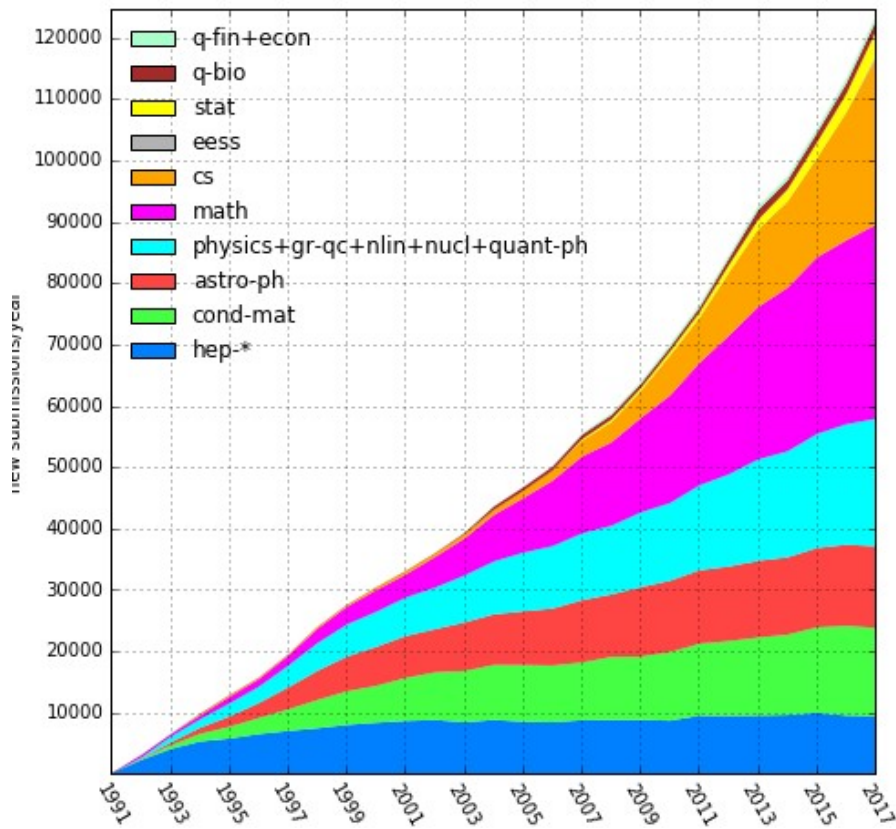
# arXivの現状

	submissions	downloads*
2012	84,000	64 M
2013	92,500	67 M
2014	97,000	90 M
2015	105,000	139 M
2016	113,380	139 M
2017	123,520	187 M



# arXiv submission rate statistics

data for 1991 through 2017, updated 31  
December 2017



47% physics, 26% math, and 22% cs.

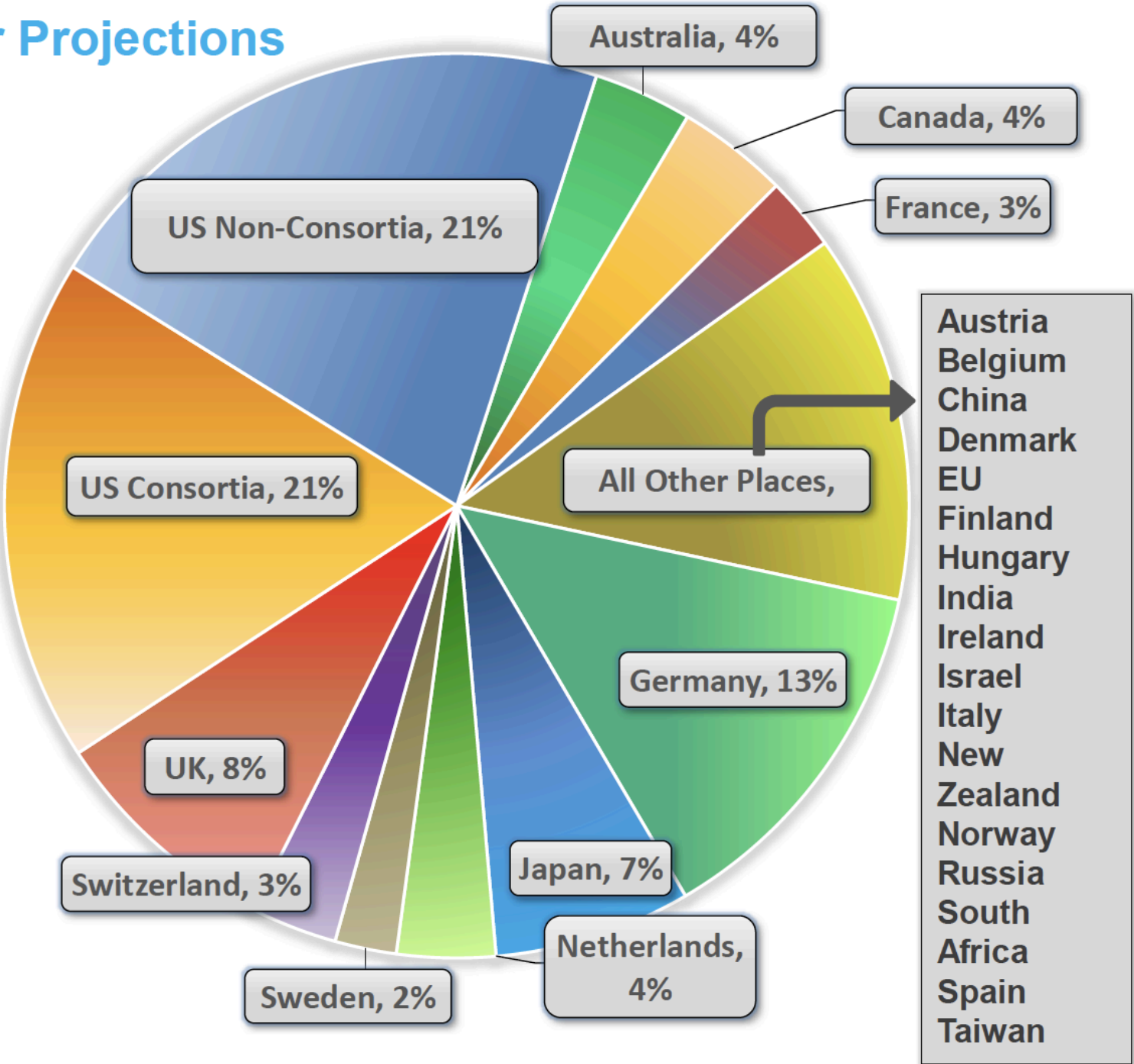
[https://arxiv.org/help/stats/2017\\_by\\_area/index](https://arxiv.org/help/stats/2017_by_area/index)

# 2018 Member Projections

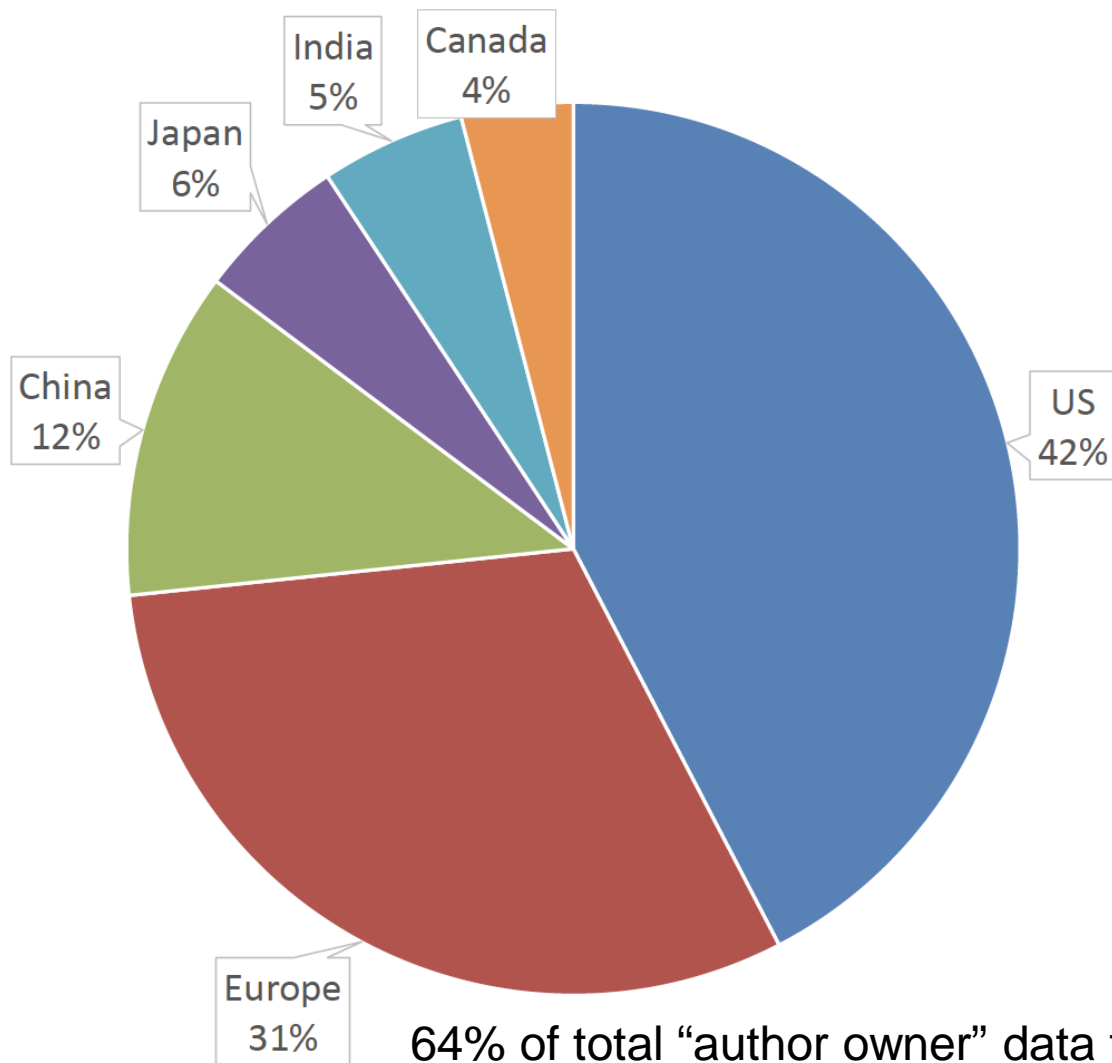
26  
Countries  
+ EU

227  
Members

\$504K  
Contributions



どこから投稿しているのか



# 運営体制

- 執行チーム:
  - Oya Rieger, Program Director, (0.5FTE)
  - Steinn Sigurdsson, Scientific Director (0.5FTE)
- 運営チーム
  - IT開発：2FTE + 4FTE
  - 運用：1FTE + 3.1 FTE
  - ユーザ対応：0.5FTE + 0.2FTE
- 技術アドバイザー 0.1FTE
- そのほか
  - Paul Ginsparg, Matte Bierbaum
  - 160名のモデレータ



# ビジネスモデル

- コーネル大学：170,000 USD + In kind
- Simons Foundation: 100,000 USD + 300,000 USD (マッチングファンド)
- メンバー機関: およそ 500,000USD
  - 機関あたり 1,000USD-4,400USD
- そのほか
- 計：1,400,000 USD程度



# arXivの現状と課題

- Preprint serverを維持し、分野を順次増やしているという面においてはarXivは順調に成長している。
- arXivが作ったpreprintの文化、およびビジネスモデルはOA時代を迎え、分野を超え、高く評価されている。
  - 多くの新興preprintはarXivをモデルに作ったと言及
- 一方、学術コミュニケーションの変貌に合わせて当初の役割と異なる役割を求められるようになった。その変化に合わせて、arXivのビジネスモデル、システムを変革できるか、これは大きな課題である。

# 台頭するpreprint サーバ

- 1991  **arXiv**: 物理学、数学、計算機科学、統計学、経済学
- 1993  **SSRN**: 社会科学 他 [Elsevier]
- 2013  **PeerJ Preprints**: バイオサイエンス [PeerJ]
- 2016  **bioRxiv**: バイオサイエンス [Cold Spring Harbor Laboratory]
- 2016  **engrxiv**: 工学 (OSF)
- 2016  **socARXIV**: 社会科学 (OSF)
- 2016  **PsyArXiv**: 心理学 (OSF)
- 2017  **BioRN**: バイオサイエンス [Elsevier]
- 2017  **LawArXiv**: 法学 (OSF)
- 2017  **AgriXiv**: 農学 (OSF)
- 2018  **ChemRxiv**: 化学[ACS, GDCh, RSC](figshare)

# プリプリントサーバの新しい役割

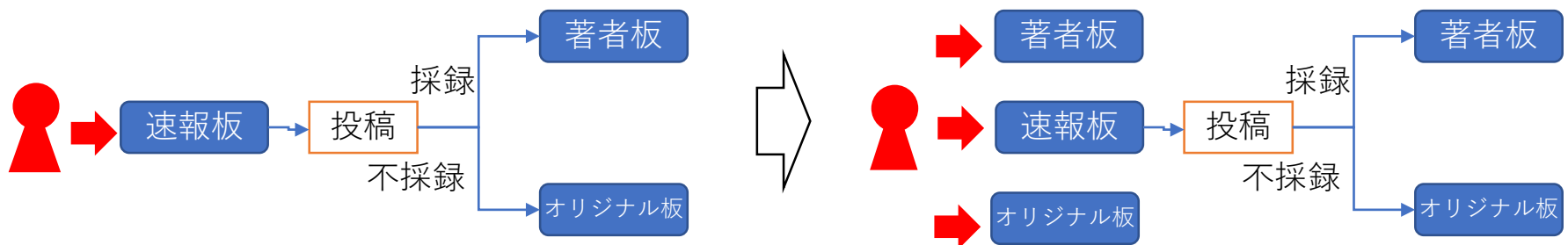
- 学術コミュニケーションのエコシステムの一部として機能することが求められている。
- すなわち、多様な出版（公開）手段の一つとしてのpreprintを位置付け、機能を再設計することが求められている。
  - 他の出版（公開）手段の役割分担
  - 他の出版（公開）手段との連携
  - オープンアクセスへ積極的な参画

# 論文出版（公開）の多様性

- ジャーナル論文だけではない多様な出版（公開）手段の発生
- プリプリントサーバ
- 機関リポジトリ
- サブジェクトリポジトリ
- オンラインプロシーディングサーバ
- オーバーレイ出版

# プリプリントサーバの役割

- 速報板
    - 発見などを素早く報告。その後、ジャーナル投稿、出版される。
  - 著者版
    - OA版をプリプリントサーバにおく
      - 出版版との共存の一つの手段
    - 速報板もジャーナル論文が出版された後は著者版の役割
  - オリジナル版
    - プリプリントサーバの論文が原板となる論文
      - 時間が勝負の論文
      - Citationされる
    - 速報板もジャーナル論文が不採録の時はオリジナル版の役割
- これまで、プリプリントサーバの役割は速報板が第一義であったが、そのプロセスで生じる著者版、オリジナル版としての利用が出てきている。



# 速報性が重視される分野での利用例

arXiv.org > stat > arXiv:1610.02920

(H)

Statistics > Machine Learning

## Generative Adversarial Nets from a Density Ratio Estimation Perspective

Masatoshi Uehara, Issei Sato, Masahiro Suzuki, Kotaro Nakayama, Yutaka Matsuo

(Submitted on 10 Oct 2016 (v1), last revised 9 Nov 2016 (this version, v2))

Generative adversarial nets from a density ratio estimation perspective  
M Uehara, I Sato, M Suzuki, K Nakayama, Y Matsuo. *arXiv preprint arXiv:1610.02920*, 2016 - arxiv.

Generative adversarial networks (GANs) are successful deep generative models. They are based on a two-player minimax game. However, the objective function derived in the original motivation is changed to obtain stronger gradients when learning the generator.

☆ 99 引用元 23 関連記事 99

## Learning in implicit generative models

S Mohamed, B Lakshminarayanan - *arXiv preprint arXiv:1610.03483*, 2016

Generative adversarial networks (GANs) provide an algorithmic framework for constructing generative models with several appealing properties: they do not require specifying a generating procedure; they provide samples from the target distribution.

☆ 99 引用元 86 関連記事 99

arXiv.org > stat > arXiv:1610.03483v1

Statistics > Machine Learning

## Learning in Implicit Generative Models

Shakir Mohamed, Balaji Lakshminarayanan

(Submitted on 11 Oct 2016 (this version), latest version 27 Feb 2017 (v4))

## Learning in Implicit Generative Models

Shakir Mohamed and Balaji Lakshminarayanan  
Google DeepMind, London  
{shakir, balajiln}@google.com

### Abstract

Generative adversarial networks (GANs) provide an algorithmic framework for constructing generative models with several appealing properties: they do not require specifying a generating procedure; they provide samples from the target distribution.

[37] M. Sugiyama, T. Kanamori, T. Suzuki, M. C. du Plessis, S. Liu, and T. Takahashi. Density-ratio estimation. *Neural Computation*, 25(10):2734–2775, 2013.

[38] M. Uehara, I. Sato, M. Suzuki, K. Nakayama, and Y. Matsuo. Generative adversarial nets from a density ratio estimation perspective. *arXiv preprint arXiv:1610.02920*, 2016.

[39] J. Zhao, M. Mathieu, and Y. LeCun. Energy-based generative adversarial network. *arXiv preprint arXiv:1609.03126*, 2016.

どちらもconference論文になってないよう。arXiv論文が直接citeされている。

# Journal, conference連携

- 最終版とのリンクがあること
- 投稿時に連携すること



# arXivはOAに貢献しない？

- Open Access and arXiv
- *But we have just learned that by far the most successful repository of open access papers in our subject, the arXiv, is to be disallowed. .. Apparently the reason they gives is that there is no mechanism for linking a paper on the arXiv with the published version of the paper.*

<https://cameroncounts.wordpress.com/2018/07/29/open-access-and-the-arxiv/>

# bioRxiv

- DOI
- bioRxivからジャーナルへの直接投稿
- 出版版との自動リンク

## PUBLICATION

bioRxiv will usually automatically add a link to the published version within several weeks of journal publication. On occasion a match is not made because the title or authors have changed or due to other contingencies. Please wait 2-3 weeks before contacting bioRxiv staff should the link not appear.

<https://www.biorxiv.org/submit-a-manuscript>

## CITATION

Preprints deposited in bioRxiv can be cited using their digital object identifier (doi).

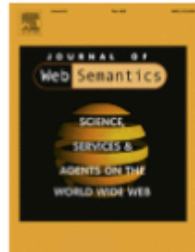
*Example:* Author AN, Author BT. 2013. My article title.  
bioRxiv doi: 10.1101/123456

## DIRECT TRANSFER FROM BIORXIV TO JOURNALS OR PEER REVIEW SERVICES (B2J)

bioRxiv can save authors time in submitting papers to journals or peer review services by transmitting their manuscript files and metadata directly from bioRxiv. This means authors do not have to spend time re-loading manuscript files and re-entering author information at the journal or peer review service website. Partners currently participating in B2J are listed below (scroll to view full list).

Acta Crystallographica Section D  
American Journal of Human Genetics  
Antimicrobial Agents and Chemotherapy  
Applied and Environmental Microbiology  
Autoimmunity Highlights  
Biochemical Journal  
Biology Open  
Biophysical Journal

<https://www.biorxiv.org/submit-a-manuscript>



# First Look Journal of Web Sem

421  
Papers

Receive Alerts

## Search Results

Search Within

The [Journal of Web Semantics](#) is an interdisciplinary journal based on research and applications of various subject areas that contribute to the development of a knowledge-intensive and intelligent service Web. These areas include: knowledge technologies, ontology, agents, databases and the semantic grid, obviously disciplines like information retrieval, language technology, human-computer interaction and knowledge discovery are of major relevance as well. All aspects of the Semantic Web development are covered. The publication of large-scale experiments and their analysis is also encouraged to clearly illustrate scenarios and methods that introduce semantics into existing Web interfaces, contents and services. The journal emphasizes the publication of papers that combine theories, methods and experiments from different subject areas in order to deliver innovative semantic methods and applications.

[Link to this page](#) | [Subscribe to this eJournal](#) (requires login)

Filter by Journal



## Journal of Web Semantics First Look

[All Journals \(421\)](#)

[JWS: Agents \(8\)](#)

[JWS: Databases & the Semantic Grid \(32\)](#)

[JWS: Human-Computer Interaction \(4\)](#)

[JWS: Information Retrieval \(25\)](#)

[JWS: Interdisciplinary Semantic Methods & Applications \(35\)](#)

[JWS: Knowledge Discovery \(2\)](#)

[JWS: Knowledge Technologies \(37\)](#)

[JWS: Language Technology \(20\)](#)

# arXivの課題

- DOI付与、メタデータの再設計 ✕
- ジャーナル、カンファレンス連携 △
- 学術コミュニケーションのエコシステムへの関与の機能強化 △
- ビジネスモデルの変革 ?
  - より頑強な財政モデル

# arXivにおける学会連携

- 学会、学会関連出版社をメンバー化
  - American Astronomical Society (AAS)
  - American Physical Society (APS)
  - American Mathematical Society (AMS)
  - Association of Computing Machinery (ACM)
  - IOP Publishing (IOPP)
  - AIP Publishing
- ACMとの連携実験
  - ACMのconferenceにおいて投稿時に論文をarXivにもおく (opt-in)

# システムの改善と arXiv Labs

- **New Generation System**への以降（順次）
  - 検索はすでに**NG**システム
- 実験的サービスの提供
  - Reference Extraction + Linking

# Learning Distributed Representations of Texts and Entities from Knowledge Base

Ikuya Yamada, Hiroyuki Shindo, Hideaki Takeda, Yoshiyasu Takefuji

(Submitted on 6 May 2017 (v1), last revised 7 Nov 2017 (this version, v3))

We describe a neural network model that jointly learns distributed representations of texts and knowledge base (KB) entities. Given a text in the KB, we train our proposed model to predict entities that are relevant to the text. Our model is designed to be generic with the ability to address various NLP tasks with ease. We train the model using a large corpus of texts and their entity annotations extracted from Wikipedia. We evaluated the model on three important NLP tasks (i.e., sentence textual similarity, entity linking, and factoid question answering) involving both unsupervised and supervised settings. As a result, we achieved state-of-the-art results on all three of these tasks. Our code and trained models are publicly available for further academic research.

Subjects: **Computation and Language (cs.CL)**; Neural and Evolutionary Computing (cs.NE)

Journal reference: Transactions of the Association for Computational Linguistics, 5 (2017), 397–411

Cite as: [arXiv:1705.02494](https://arxiv.org/abs/1705.02494) [cs.CL]  
(or [arXiv:1705.02494v3](https://arxiv.org/abs/1705.02494v3) [cs.CL] for this version)




## Bibliographic data



Select data provider: **Semantic Scholar** [\[Disable Bibex \(What is Bibex?\)\]](#)

### References (46)

Data provided by:  Semantic Scholar  
[\(report data issues\)](#)

Filter:

Sort: Influence   

Pages:  1 2 3 4 5  Skip: 1 


A Neural Network for Factoid Question Answering over Paragraphs *EMNLP 2014*  
Mohit Iyyer, Jordan L. Boyd-Graber, Leonardo Max Batista Claudino, Richard Socher, Hal Daumé






DBpedia Abstracts: A Large-Scale, Open, Multilingual NLP Training Corpus *LREC 2016*  
Martin Brümmer, Milan Dojchinovski, Sebastian Hellmann




### Citations (14)

Data provided by:  Semantic Scholar  
[\(report data issues\)](#)

Filter:

Sort: Influence   

Pages:  1 2  Skip: 1 

 Improving Entity Linking by Modeling Latent Relations between Mentions *ArXiv 2018*  
Phong Le, Ivan Titov



Neural Collective Entity Linking *COLING 2018*  
Yixin Cao, Lei Hou, Juan-Zi Li, Zhiyuan Liu



RPI BLENDER TAC-KBP2017 13 Languages EDL System *TAC 2017*

# なぜ、研究者は論文（書籍）を書くのか？

- 記録を残したい
- 人に知らせめたい
- コミュニティで評価されたい
- 仕事を獲得・維持したい



# なぜ、研究者は論文（書籍）を書くのか？

- 記録を残したい

- 保存性(preservable)

- 人に知らしめたい

- アクセス性(accessible)

- 識別性(indetifiable)

- コミュニティで評価されたい

- 発見性(discoverable)

- 社会性(socially embedded)

- 指標性(indexed)

- 仕事を獲得・維持したい

# なぜ、研究者は論文（書籍）を書くのか？

往復書簡、未刊行書籍

- |                 |   |
|-----------------|---|
| • 記録を残したい       | • 保存性(preservable)  |
| • 人に知らせめたい      | • アクセス性(accessible)<br>• 識別性(indetifiable)                        |
| • コミュニティで評価されたい | • 発見性(discoverable)<br>• 社会性(socially embedded)<br>• 指標性(indexed) |
| • 仕事を獲得・維持したい   |   |

# なぜ、研究者は論文（書籍）を書くのか？

## 書籍出版

- 記録を残したい

- 保存性(preservable)

(図書館)

- 人に知らしめたい

- アクセス性(accessible)

- 識別性(indetifiable)

- コミュニティで評価されたい

- 発見性(discoverable)

- 社会性(socially embedded)

- 指標性(indexed) (書評)

- 仕事を獲得・維持したい

# なぜ、研究者は論文（書籍）を書くのか？

## ジャーナル論文出版

- |                 |   |
|-----------------|---|
| • 記録を残したい       | • 保存性(preservable)<br>(図書館)   |
| • 人に知らしめたい      | • アクセス性(accessible)<br>• 識別性(indetifiable)  |
| • コミュニティで評価されたい | • 発見性(discoverable)<br>• 社会性(socially embedded)<br>• 指標性(indexed)<br>(citation index) |
| • 仕事を獲得・維持したい   |   |

# なぜ、研究者は論文（書籍）を書くのか？

## 紀要、Technical Report

- 
- |                 |                          |
|-----------------|--------------------------|
| • 記録を残したい       | • 保存性(preservable)       |
| • 人に知らしめたい      | • アクセシビリティ(accessible)   |
| • コミュニティで評価されたい | • 識別性(indetifiable)      |
|                 | • 発見性(discoverable)      |
|                 | • 社会性 socially embedded) |
|                 | • 指標性(indexed)           |
| • 仕事を獲得・維持したい   |                          |

# なぜ、研究者は論文（書籍）を書くのか？

電子ジャーナル論文出版

(CLOCKSS)

- 記録を残したい

- 保存性(preservable)

- 人に知らせめたい

- アクセス性(accessible)

- 識別性(indetifiable)

(DOI)

- コミュニティで評価されたい

- 発見性(discoverable)

- 社会性(socially embedded)

- 指標性(indexed)

(citation index)

- 仕事を獲得・維持したい

# なぜ、研究者は論文（書籍）を書くのか？

## プリプリント出版

- 記録を残したい
- 保存性(preservable)
- 人に知らせめたい
- アクセス性(accessible)
- 識別性(indetifiable)
- コミュニティで評価されたい
- 発見性(discoverable)
- 社会性 socially embedded)
- 指標性(indexed)
- 仕事を獲得・維持したい

# なぜ、研究者は論文（書籍）を書くのか？

## プリプリント出版

- 記録を残したい
- 保存性(preservable)
- 人に知らせめたい
- アクセス性(accessible)
- 識別性(indetifiable)
- コミュニティで評価されたい
- 発見性(discoverable)
- 社会性(socially embedded)
- 電子ジャーナル論文出版 指標性(indexed)

- 仕事を獲得・維持したい



# プリプリントサーバの役割

## ジャーナルとの連携

プリプリント出版

- |                 |   |
|-----------------|---|
| • 記録を残したい       | • 保存性(preservable)                              |
| • 人に知らせめたい      | • アクセス性(accessible)<br>• 識別性(indetifiable)      |
| • コミュニティで評価されたい | • 発見性(discoverable)<br>• 社会性(socially embedded) |
| 電子ジャーナル論文出版     |   |
| • 仕事を獲得・維持したい   | • 指標性(indexed)                                  |



ジャーナルとの機能分担（オープンアクセス、速報性）

# プリプリントサーバの役割 オリジナル論文

プリプリント出版

- 
- The diagram consists of a large rounded rectangle with a red border, divided into four horizontal colored bands. The top two bands (orange and yellow) are grouped by a blue border. The bottom two bands (light blue and light green) are also grouped by a blue border. A large blue arrow points from the yellow band down to the light blue band. The text is organized into two columns within each band.
- 記録を残したい
  - 人に知らせめたい
  - コミュニティで評価されたい
  - 仕事を獲得・維持したい
  - 保存性(preservable)
  - アクセス性(accessible)
  - 識別性(indetifiable)
  - 発見性(discoverable)
  - 社会性(socially embedded)
  - 指標性(indexed)

# プリプリントサーバの役割

## プリプリント出版

- 記録を残したい
- 保存性(preservable)
- 人に知らせめたい
- アクセス性(accessible)
- 識別性(indetifiable)
- コミュニティで評価されたい
- 発見性(discoverable)
- 社会性(socially embedded)
- 電子ジャーナル論文出版指標性(indexed)
- 仕事を獲得・維持したい

# まとめ

- **Preprint**サーバは学術コミュニケーションのエコシステムの重要なパートになるようになった。
  - 速報性のための出版手段
  - オリジナル出版
  - オープンアクセス出版（著者版）
  - ジャーナル連携
- そのために求められる機能が増えてきている
  - DOI
  - 出版版のリンク
  - アーカイブ
- **arXiv**はこの変革に対して、システムの機能、マネジメント、ビジネスモデルそれぞれにおいて対応が遅れ気味である。
- しかし、**arXiv**は後発の多くのプリプリントサーバのロールモデルであり、また**arXiv**も着実に分野・投稿量・アクセス量を増やしており、一層の発展が期待される。