

オープンサイエンスとデータ駆動型科学：  
データのサイエンスが目指すものは何か  
(試論)

村山泰啓  
(NICT)

# 科学とデータ？

オープンサイエンス

もうすこし考え方を整理したい

データ  
サイエンス

データ駆動型科学、  
データ中心科学

# Open Access to Open Data and Open Science

## Overview example

[図版は林和弘氏（文部科学省科学技術・学術政策研究所上席研究官）による。(2015)]

Database Repository

Data Sharing

Open Research Data

Data journal

Self Archiving

Institutional Repository

Full OA (mega) journal

Science Commons

Science 2.0+

(Scholarly activity)

Open Access

Open Science

(Common activity)

Article

Access

Research Outputs

ReUse

Research Activity

Open Source

Open Innovation

Citizen Science

Creative Commons

Open Data  
(Open gov)

Code for XX

2000's

Improve,  
Incremental

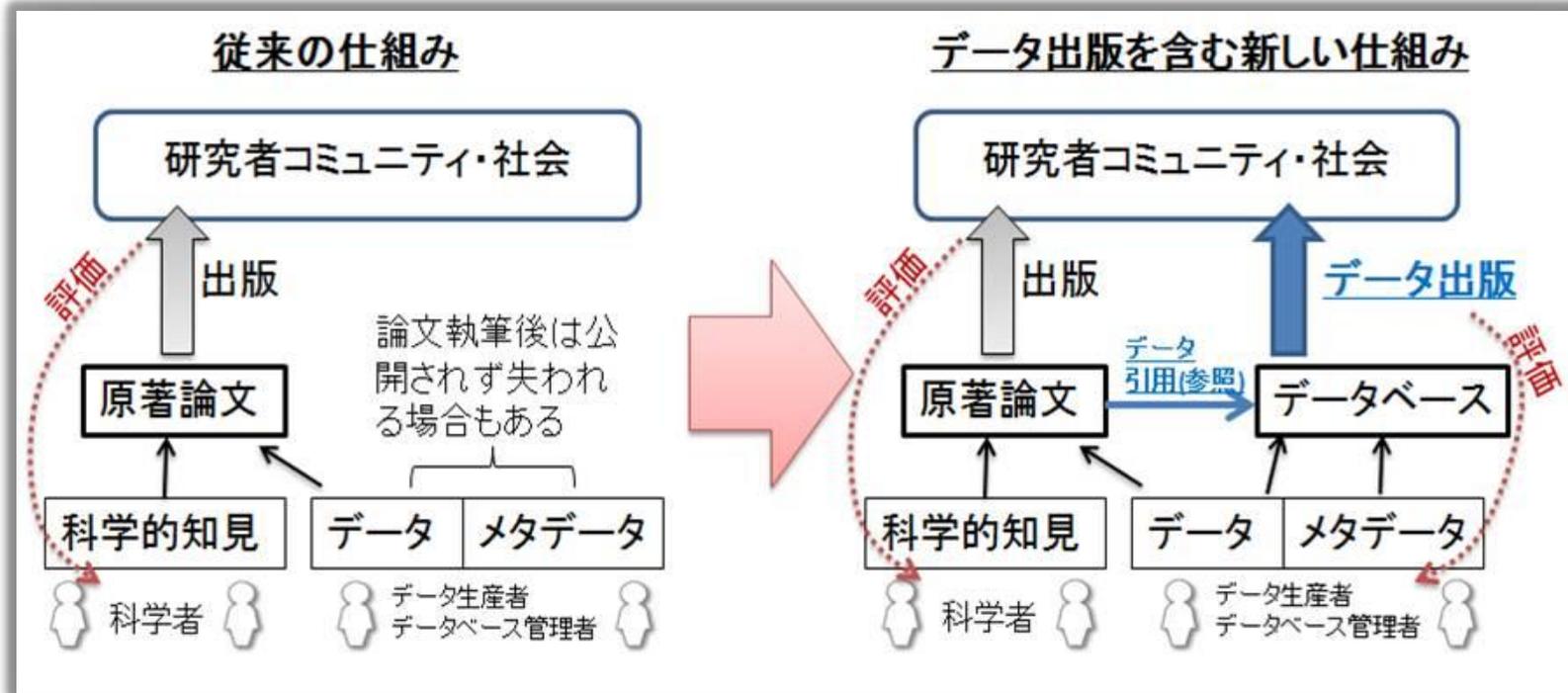
2010's

Redesign,  
Disruptive?

# 論文だけでなく、データも研究成果

## ■ データを公表(出版)・引用・参照すると

- 論文・書籍と同様、知的生産力の基準に。→ 研究職・教育職の業績評価。
- 信頼できるデータの生成・提供は現代では科学者の仕事ではないか。← 評価



## ■ データパブリケーション

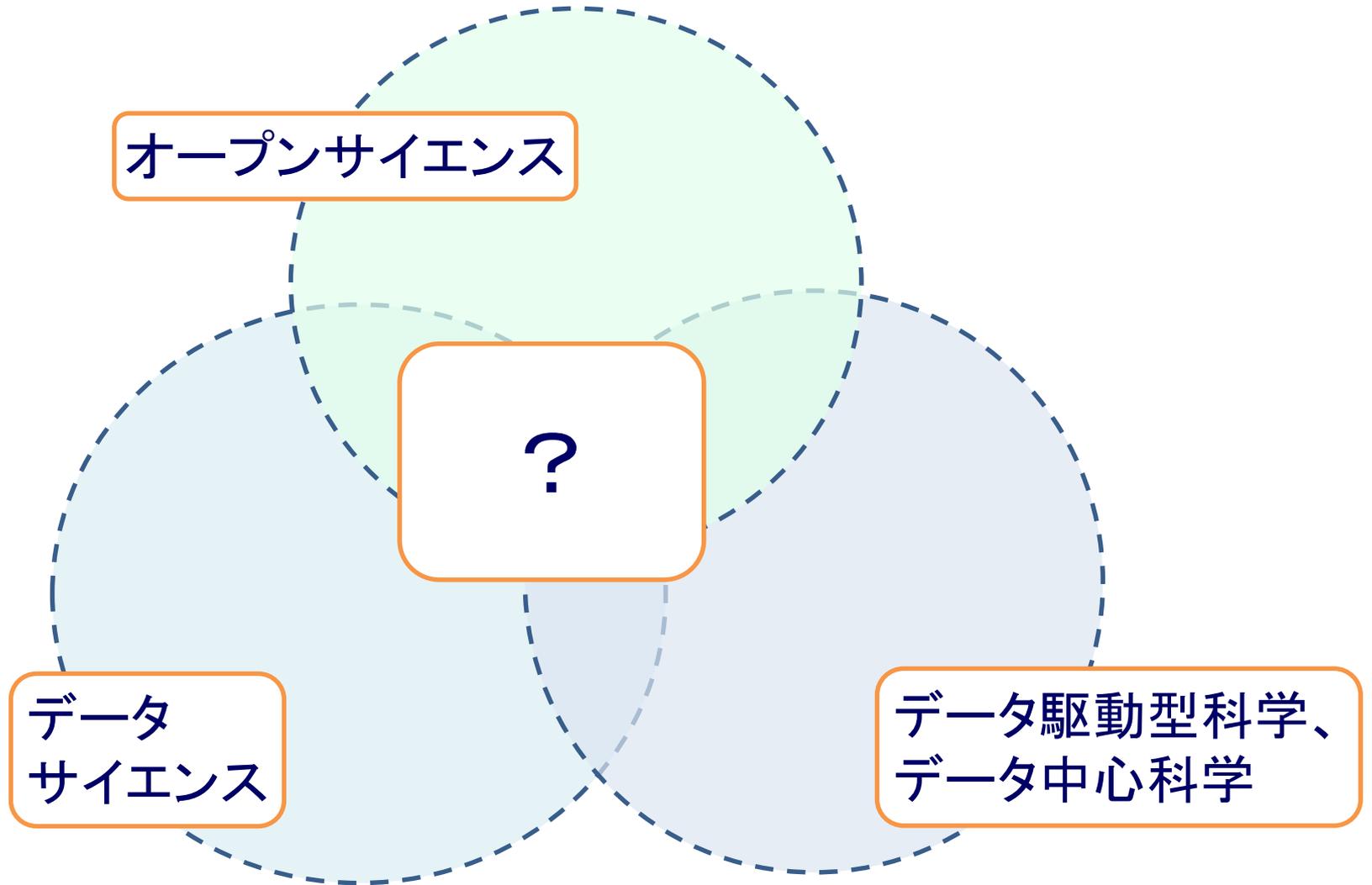
[地球電磁気・地球惑星圏学会, 2013]

- 課題: データの「査読」「固定」「公表」等をどうするか。
- 課題: ID標準化、引用ルール確立、評価手法など国際団体等で模索中

## ■ データサイテーション

- データを文献のように「引用」「参照」する仕組み
- 課題: ID標準化、引用ルール確立、評価手法など国際団体等で模索中

# 科学とデータ？



# データと科学

- データ集約型科学、データ中心科学、データ駆動型科学
  - 「第4のパラダイム」: 実験科学、理論科学、計算科学、の次に来る新たな科学の形
  - データから新たな科学的発見を、新たな科学的価値を。
- データサイエンス
  - データに関する科学。多くの・多様なデータを活用して新たな価値を生み出す研究分野( 関西大学データサイエンス研究センター<http://www2.kansai-u.ac.jp/dslab/datascience.html>, 2018/9/13 accessed )
  - データの分析についての学問分野。統計学、数学、計算機科学などに関連し、主に大量のデータから、何らかの意味のある情報、法則、関連性などを導き出すこと、またはその処理の手法に関する研究を行う。(デジタル大辞泉(小学館))
  - データ中心科学、データ駆動型研究と同義で語られる場合も散見される
- オープンサイエンス
  - ネット上集合知、市民科学
  - デジタルデータ流通と学術規範、創造性の変革

- データサイエンスとデータ駆動型科学:

「. . . 大学等におけるデータ駆動型学術研究(データサイエンス)の推進を. . . 」(2018年度 情報・システム研究機構 データサイエンス共同利用基盤施設「ROIS-DS-JOINT」募集要項)

- オープンサイエンスの定義がさだまらないこと:

ベルリンで行われたオープンサイエンス会議2017において、. . . . オープンサイエンスがどういうものかについて、実に多種多様な考え方がみられた。 . . . T定義することが時期尚早などというより. . . オープンサイエンスの定義が1つに絞れないというべきで. . . (Defining Open

Science Definition, 2017, <https://im2punt0.wordpress.com/2017/03/27/defining-open-science-definitions/>, 2018/9/13 accessed)

# 政策的オープンサイエンス文脈での科学研究システムの拡張

研究 ≠ 科学

社会への貢献

研究活動

Conventional Science Method

Question

Hypothesis

Experiment/  
Observation

Analysis/  
Discussion

Conclusion

Facility

Software

Data  
Contribution

Data

Paper

Author

Reviewer

Affiliation

Editor

Funder

Publisher

General Society/  
political decision making  
一般社会・政治的意思決定

Community Consensus  
Forming  
科学コミュニティのコンセンサス

Open discussion

近年よく議論される  
「オープンサイエンスっぽい」  
領域

再検証  
search

コミュニティでの研究情報共有

コミュニティでの共有



# オープンサイエンスの定義例

- 欧州委員会:

オープンサイエンスは、デジタルテクノロジーやを用いた相互協力および知識の伝搬に基づく、科学研究への新しいアプローチといえる。

(European Commission (2016), Open innovation, Open Science, open to the world. A vision for Europe, 2016b:DOI: 10.2777/061652.).

- OECD:

オープンサイエンスは公的資金による主な研究成果(論文出版と研究データ)を一般にたいして、最低限の制約をもってデジタルフォーマットでアクセス可能とすること

(OECD (2015), Making Open Science a Reality, <http://dx.doi.org/10.1787/5jrs2f963zs1-en>)

「EU、OECD型のオープンサイエンス」?

「トップイニシアチブによるオープンサイエンス」?

将来的には: デジタル社会基盤の構築へつながるオープンサイエンス?

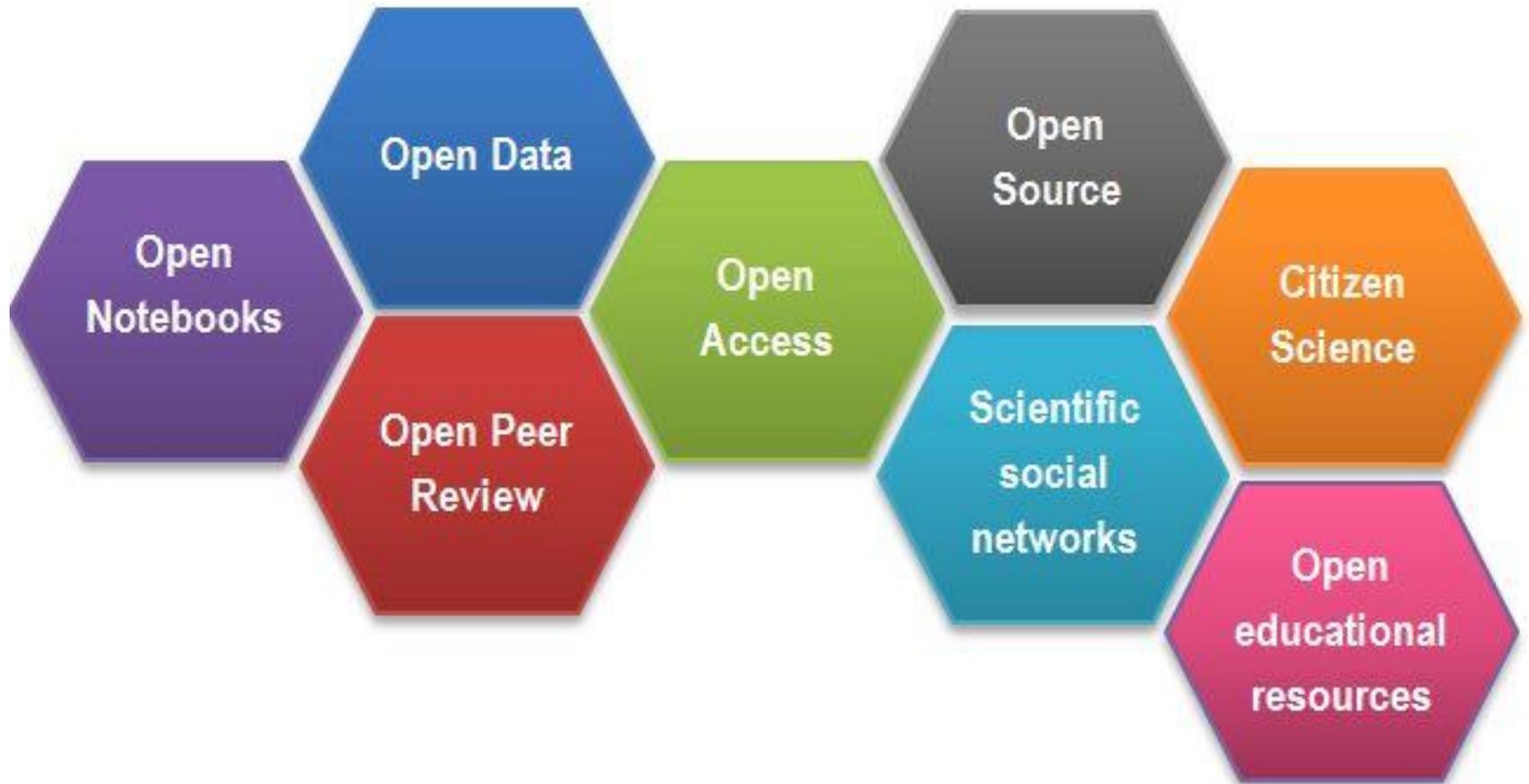
# オープンサイエンス革命

- オープンサイエンス革命 マイケル・ニールセン著 (日経(2013)書評記事より)
- 市民参加による科学探究を提言
- 科学者がネット上で広く情報を積極的に開示、衆知を集め問題を解く「オープンサイエンス」



# オープンサイエンスのさまざまな切り口

EC Horizon2020プロジェクト「FOSTERプロジェクト」(2014～)



# データサイエンスとは？

統計学？

機械学習？

データに関する研究を行う学問

データサイエンスの  
明確な定義はない



# “Greater Data Science” (GDS)

- Donoho (2017)で言及(おそらくはじめて)。
- 統計学分野で議論されてきた「データサイエンス」の再考に狙いがあるらしい。
- Ver.1は2015年に著者の個人サイトで公開された模様。
- 参考文献 : David Donoho (2017) 50 Years of Data Science, Journal of Computational and Graphical Statistics, 26:4, 745-766,  
doi:10.1080/10618600.2017.1384734

# David Donoho



<https://profiles.stanford.edu/david-donoho>

- スタンフォード大学数学研究科・教授
- プリンストン大学卒業 (John Tukey (1915-2000) に師事)、ハーバード大学・博士学位取得 (1983)
- 統計理論 (「マルチスケール幾何解析」や「スペース・サンプリング (疎な標本化) 理論」等) を研究。
- マッカーサーフェロー (1991)、フォン・ノイマン賞 (2001)、Gauss Prize (2018)、米数学学会 (AMS) フェロー、等受賞

# Donoho (2017)より

(From U.-Michigan Data Science Initiative Web, accessed 2015)

- データサイエンスは第4のアプローチ

*“Data science has become a fourth approach to scientific discovery, in addition to experimentation, modeling, and computation,” said Provost Martha Pollack.*

- 科学的発見と実践の結合には、多様で大量のデータの収集、マネジメント、処理、解析、可視化、解釈がふくまれる

*“This coupling of scientific discovery and practice involves the collection, management, processing, analysis, visualization, and interpretation of vast amounts of heterogeneous data associated with a diverse array of scientific, translational, and inter-disciplinary applications.”*

# GDS (Greater Data Science)

[Donoho, 2017]

すべてが、データサイエンス

下積み・  
雑用??

1. *Data Gathering, Preparation, and Exploration*  
データ収集、データ整備、データチェック

下積み・  
雑用??

2. *Data Representation and Transformation*  
データ・DBの形式整備、再構成

つながり  
一般的には論文に  
そうなる仕事

3. *Computing with Data*  
データ処理

4. *Data Modeling*  
データ生成・予測のモデル(?)

5. *Data Visualization and Presentation*  
データの可視化(通常 of データ描画 ~ 高度化)

下積み・  
雑用??

6. *Science about Data Science*  
データサイエンスに関するサイエンス

1. *Data Gathering, Preparation, and Exploration*

データ収集、データ整備、データチェック

2. *Data Representation and Transformation*

データ・DBの形式整備、再構成

データ収集

データキュレーション

アノテーション

メタデータ生成

FAIR data化

6. *Science about Data Science*

データサイエンスに関するサイエンス

GDSにおける

方法論、理論、手法

等のサイエンス

と思われる

# (デジタルサイエンス視点から見た) オープンサイエンスの構成要素例

研究施設

ソフトウェア

論文発表の場(学術  
ジャーナル、カン  
ファレンス等)

データ共有,  
FAIRデータ利用  
(Findable, Accessible,  
Interoperable,  
Reusable)

予算機関、  
政府

学協会

研究機関

市民・社会との  
相互理解・共創

オープンソース

ツール共有

OAジャーナル

ジャーナルポリシー  
(データ引用等)

Green OA/Gold OA

Subscription model

研究データ  
オープン化

研究データ保存

データ情報の組化  
標準化、

データ引用・参照

データ識別子、メタデータ

(DMP)

データマネジメントプラン

オープン化推進力等

公的データ基盤の整備

文化、慣習、研究者評価

文献・機関リポジトリ

データ・機関リポジトリ

研究者識別子  
(国際・横断的)

研究業績評価

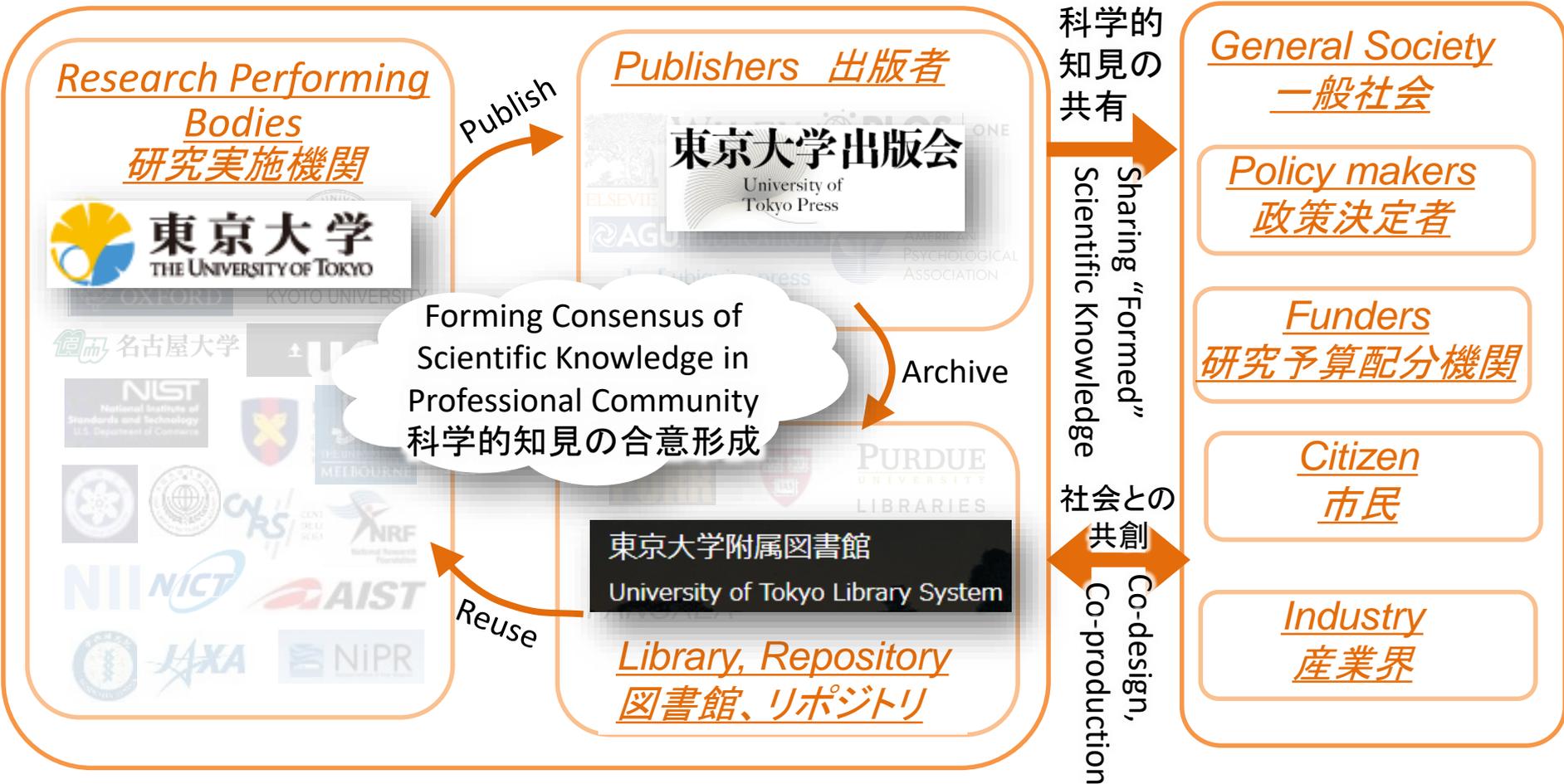
シチズンサイエンス

オープンサイエンス

オープンイノベーション

# Science as a Social System (on "Print" technology so far)

## 社会システムとしてのサイエンス (これまでは印刷文化・技術)



# まとめにかえて

社会と科学の関わり方、相互信頼・相互利用、  
co-design, co-production

データ  
サイエンス、  
データ駆動型  
科学

オープンサイエンス

さらに広い「Greater Data Science」