

学術コミュニケーションの動向

2017.7.13

佐藤義則

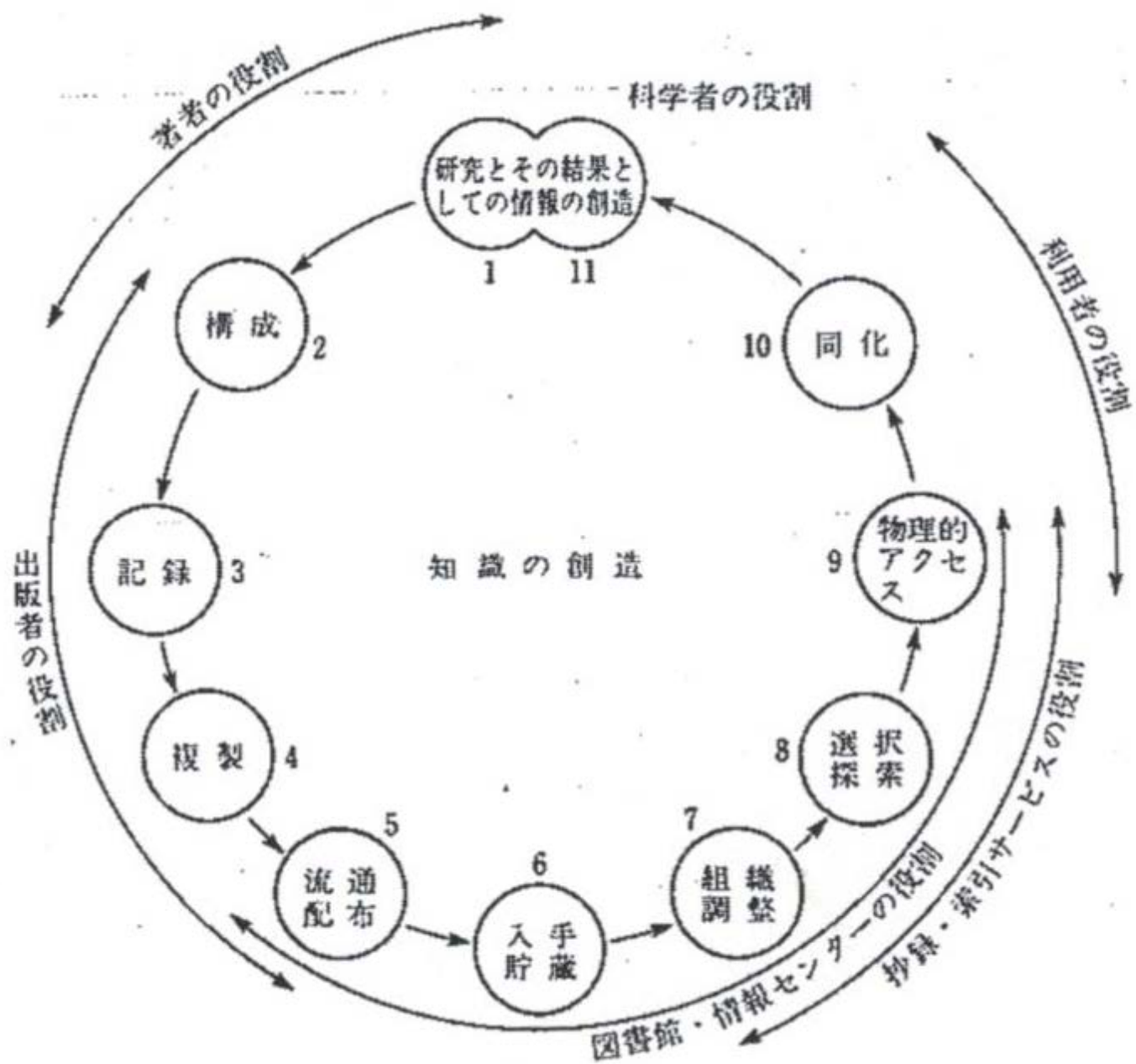
東北学院大学

デジタルコンテンツの繁栄は、見方によっては、
学術コミュニケーションの進化であり、革命であ
り、あるいは危機である。

- Borgman, Christine L. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, Massachusetts; MIT Press, 2007, p. 9.

学術コミュニケーションとは

- 「研究その他の学術的著作が生み出され，品質を評価され，学術コミュニティに広められ，そして将来の利用のために保存されるシステム」
 - Association of Research Libraries. “Scholarly Communication.” http://www.arl.org/focus-areas/scholarly-communication#.U4GZupR_vUQ



「学術情報の還流過程」(図書館情報学ハンドブック. 第2版, 1988.3, p. 256)

研究プロセス全般への関与 (パデュー大学図書館の例)

- ❖ 学術コミュニケーションとは、教育、研究、学問に関わる知識の創成から、流通、保存に至るまでの、研究を遂行し結果を共有するプロセスのことである。
- ❖ パデュー(大学)の図書館は伝統的に、他のたいていの学術図書館と同様に、図書、雑誌、その他の情報資源を収集し、現在および将来の学生、研究者、学者のためのコレクションを構築してきた。しかし、パデューの利用者は、学術コミュニケーション・サイクルの最初(例えば、データ管理やデータ・キュレーション)から、出版(公開)や流通(オープンアクセス・リポジトリや雑誌)を通じて、学術コミュニケーションに関連する追加のサービスを必要としており、パデューの図書館はこうした領域で支援を行うことができる。

学術コミュニケーションの変化

- 変化の要因

- 技術の進展

- デジタル技術, インターネット, 電子ジャーナル, SNS...

- 研究者

- 図書館

- 政府機関(および資金提供団体)

※技術的要因とともに, 政治的, 社会経済的要因

学術コミュニケーションの変化と 利害関係者 (stakeholders)

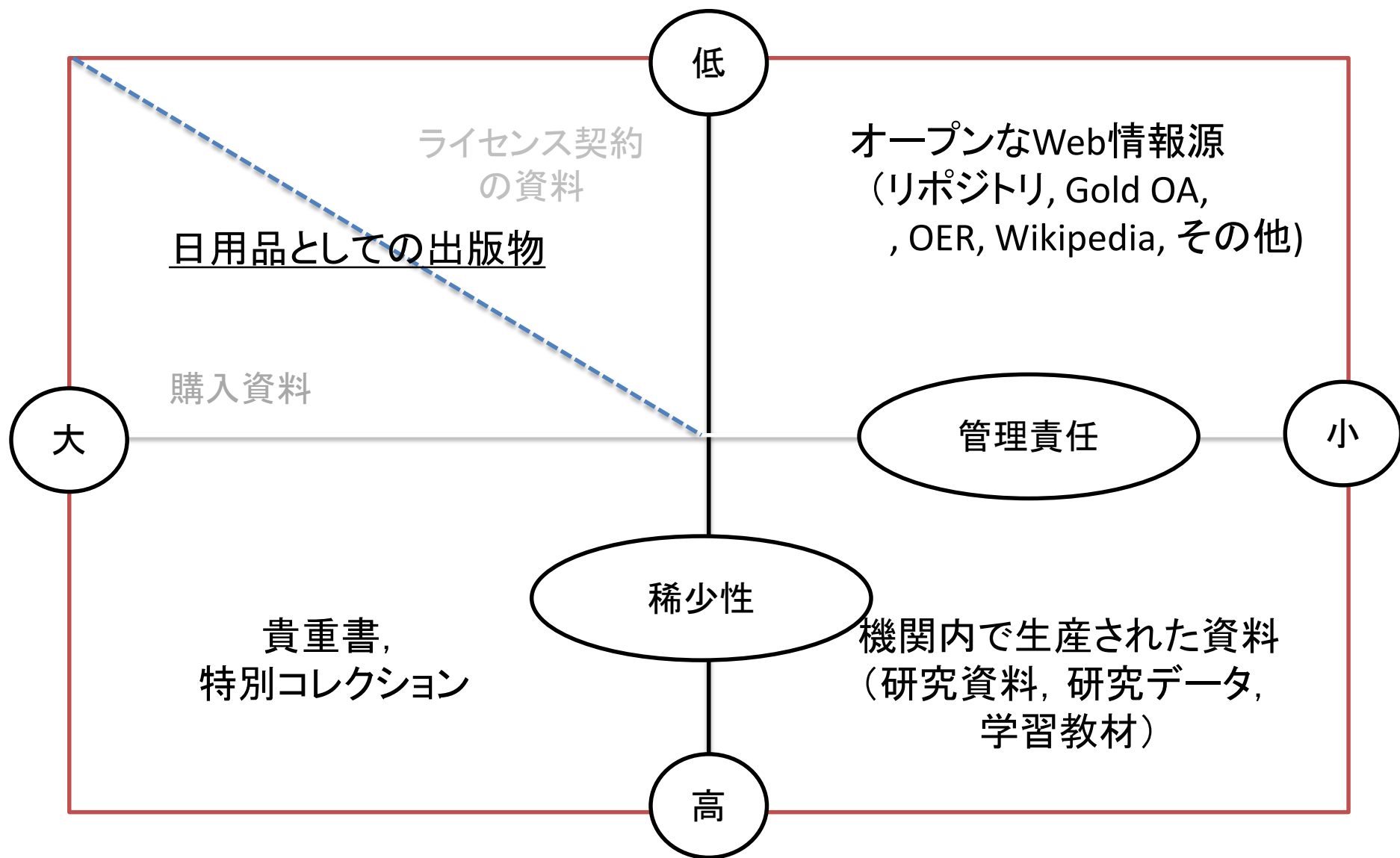
- 研究者
 - 量的変化
 - 新たな研究分野の出現, 研究分野の細分化
 - 研究者数の増大 (特に, 新興国)
 - 質的变化
 - デジタルな研究環境への進行に伴う一連の活動スタイルの変化
 - 検索, 収集, リーディング, 執筆, 連携, 観察, ノート作成, 翻訳, データ採取等におけるデジタル機器の利用
 - 学術的記録 (“Scholarly Record”) の範囲の広がりおよび量的拡大
- 出版者
 - 1960年代以降の論文量の増大と巨大産業への成長
 - 1980年代以降のM&A (合併・統合) の進行と寡占化
 - 電子ジャーナルおよびビッグディール
- 図書館
 - 電子図書館
 - 機関リポジトリとオープンアクセス
- 政府機関 (および資金提供団体)
 - オープンアクセスの義務化 (研究資金の効率的運用)

デジタル化による変化

- バックランド 『図書館サービスの再構築』勁草書房, 1994
 - 場所的制約を受けない(どこからでも)
 - 複数の人々が同時に利用可能
 - 容易に複製可能
 - 柔軟性に富む(改訂、再編成、再初期化、結合などが容易)
 - 場所をとらない
- 加えて,
 - 多様な情報形式の混在が可能(テキスト、画像(写真、ビデオ)、音声)
 - 多様な粒度
 - オブジェクト間のリンク, データ間のリンク
 - 検索が容易; 発見可能性の向上(見えないものは, 無いものと同じ)
 - 流通が容易(むしろ, 流通を制限するのが課題)
 - 利用記録(捕捉できなかったものが捕捉可能に, use metadata)

図書館コレクションの変化

- 館内所蔵から，ウェブ上の情報資源へ
（所蔵からアクセスへ）
- 単独のコレクションから，集合的コレクションへ
（collective collection）
- 共有資源から，公共財へ
- 課題
 - － 効率性，持続可能性，公平性の確保
 - － コスト負担（誰が，どのように）
 - － 保存問題，および真正性の維持



出典: Malpas, Constance. "Scarcity and Abundance: the Cooperative Imperative in Special Collections," 53rd Annual RBMS Preconference, 20 June 2012.
available at <http://www.oclc.org/research/presentations/default.htm> (2015-09-01 確認)

流通市場とライセンスリング

- シリアルズクライシスの背景（Bergstrom & Rubinfeld, 2010）
 - 低い価格弾力性（price elasticity）
 - 価格が上昇しても、図書館あるいは大学は購読を継続
 - 二面性の市場（two-sided market）
 - 学術雑誌市場の、出版社 — 研究者（＝著者）、出版者 — 図書館（購買契約）、図書館 — 研究者（＝利用者）という構図（Bergstrom & Rubinfeld, 2010）

※ 結果としての、出版社間の競争あるいは新規参入が生じにくい構造

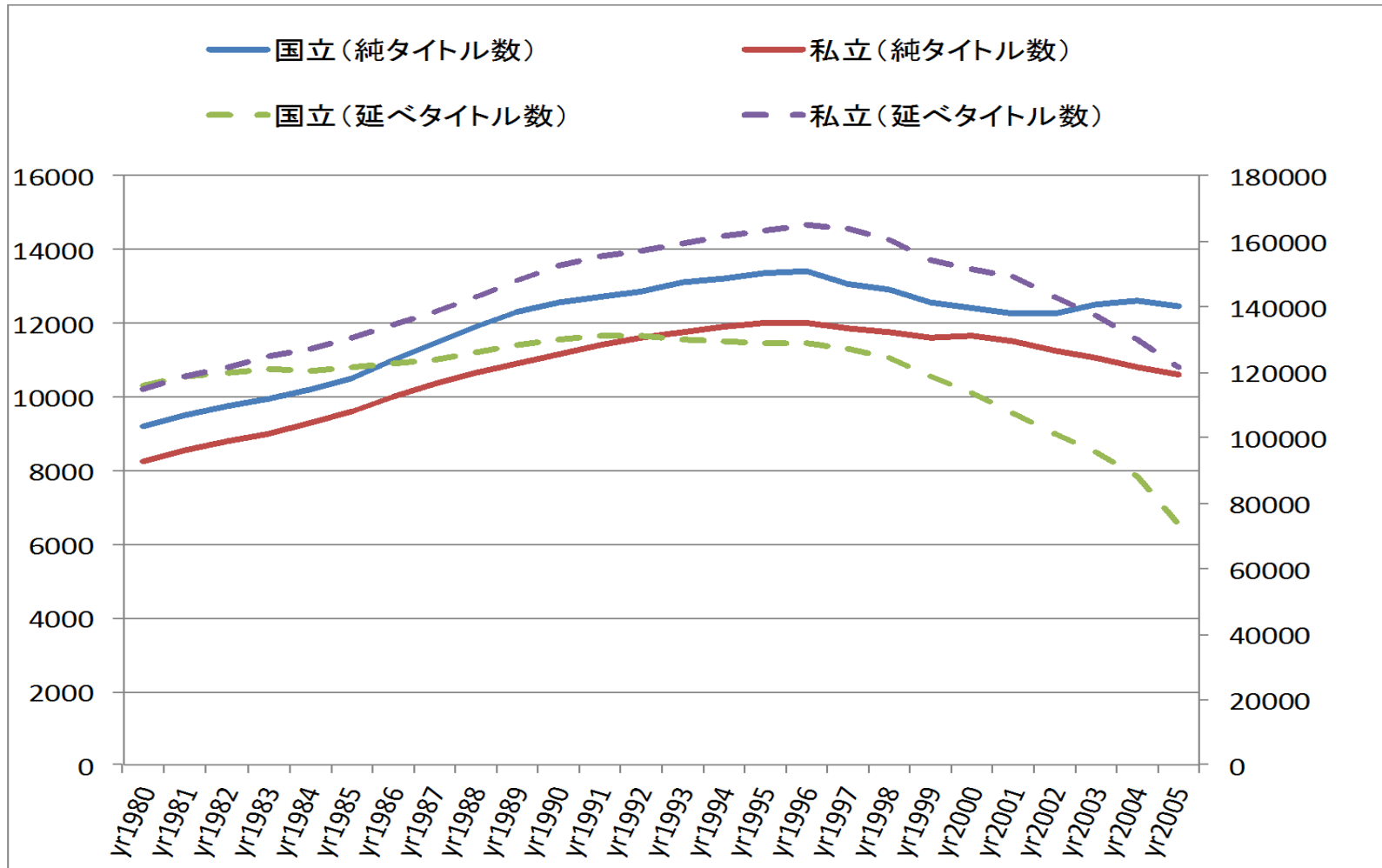
二面性の市場 (Two-sided Market)

- プラットフォームとは、生産者と利用者の間の相互作用において「自明ではない」役割を果たす仲介手段

買い手	プラットフォーム	売り手
ゲーマー	ビデオゲーム・プラットフォーム	ゲーム開発者
利用者	オペレーティング・システム	アプリ開発者
視聴者, 利用者	新聞, テレビ, 検索エンジン等	広告スポンサー
カード保有者	クレジットカード	小売業者
図書館	学術雑誌	論文の著者

- プラットフォーム市場の特徴: ネットワーク外部性(いずれの側の参加者も, もう一方の側からの参加者が増えればより高く評価するようになり, そして多くの場合自らの側についても同様)
- 比較的少数の競合プラットフォームで構成されるのが一般的で, 二つの市場のうち少なくとも一つで強い独占力を形成

日本におけるシリアルズクライシス (～1990年代)

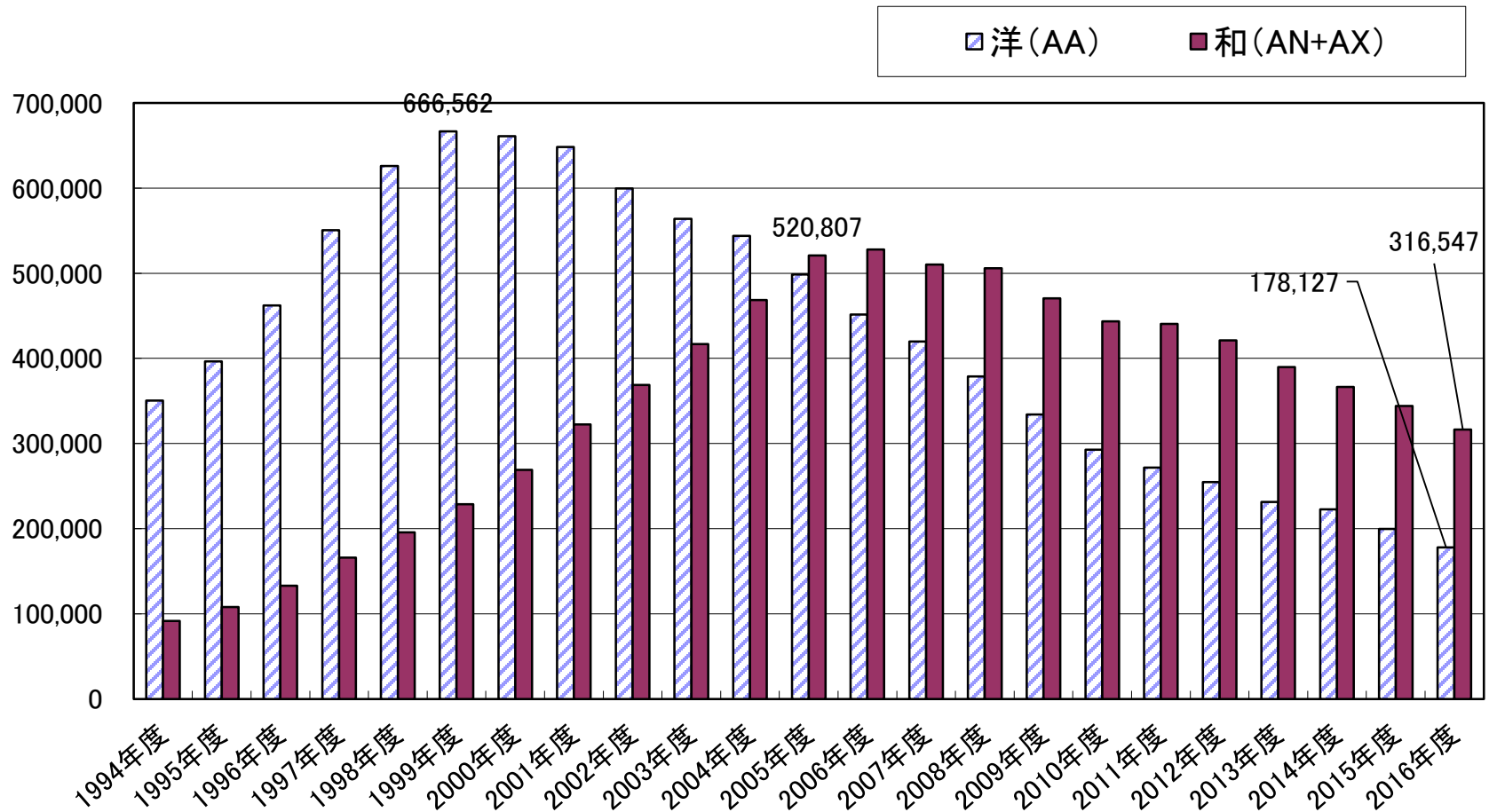


NACSIS-CATにおける英文の外国雑誌タイトル数の変化

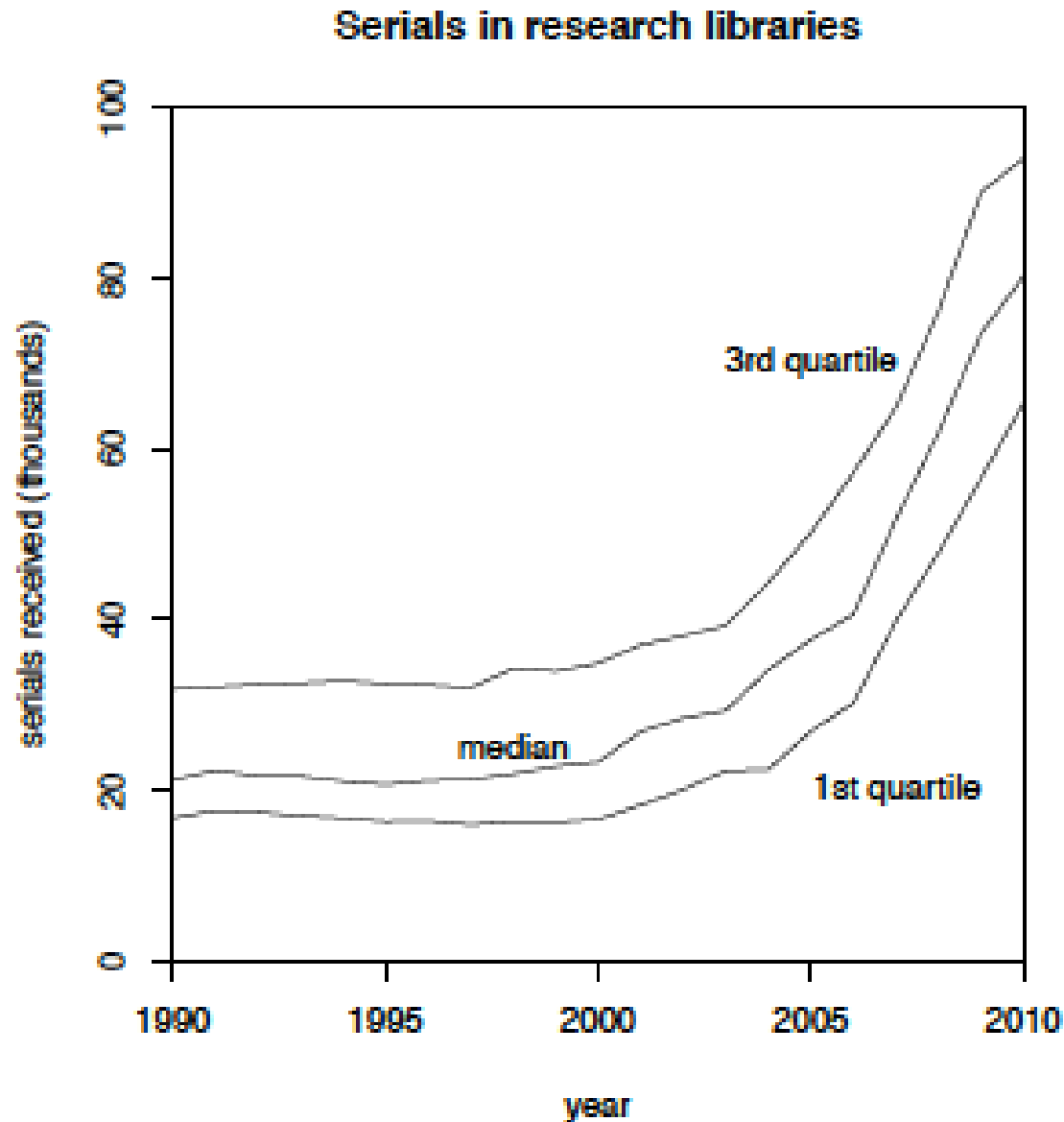
ビッグディールの影響

- 学術雑誌の利用環境の向上
 - いつでも, どこからでも(研究室 + 学認 or vpn利用), 従来よりも多くのタイトルへアクセス
- 図書館サービスの構図の変化
 - 例. ILLの減少(オンラインアクセス環境の向上の間接的表現)
 - 洋雑誌掲載論文への複写依頼は約20万件で、前年度比約 $\Delta 9.2\%$ (昨年度は約 $\Delta 3.8\%$)。ピークは1999年度の約67万
 - 和雑誌掲載論文への複写依頼は約34万件で、前年度比約 $\Delta 6.1\%$ (昨年度は約 $\Delta 6.0\%$)。ピークは2006年度の約53万件
- 利用者の意識の変化
 - 例. SCREAL調査における「印刷体雑誌の必要性」。利用者の意識はサービスの普及や受容に合わせて変化している

ILLの減少 (オンラインアクセス環境向上の効果)



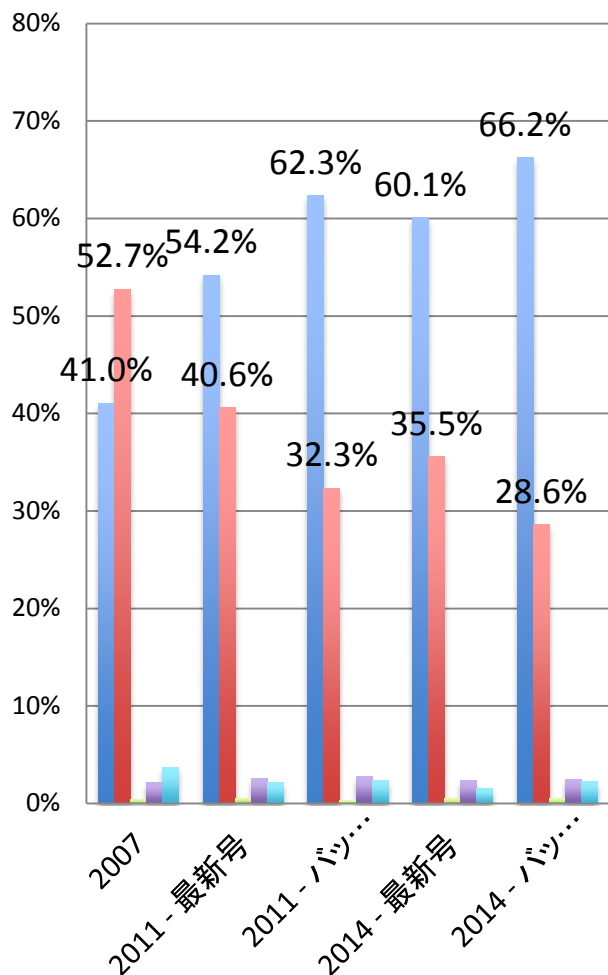
NACSIS-ILL複写リクエスト(完了分)の和洋別推移



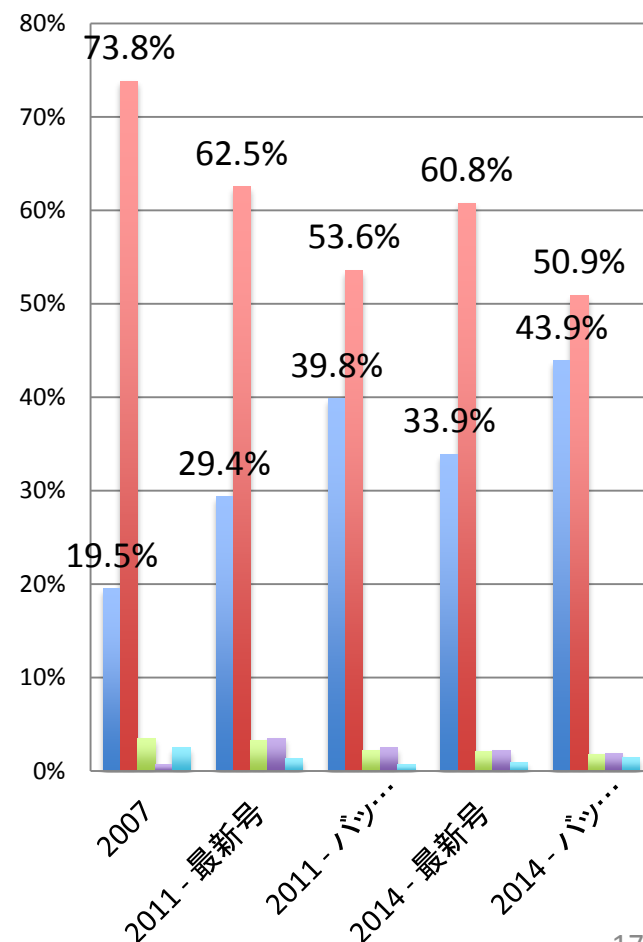
Andrew Odlyzko, "Open Access, library and publisher competition, and the evolution of general commerce." [arXiv:1302.1105 \[cs.DL\]](https://arxiv.org/abs/1302.1105)

印刷体雑誌の必要性(2007 - 2014)

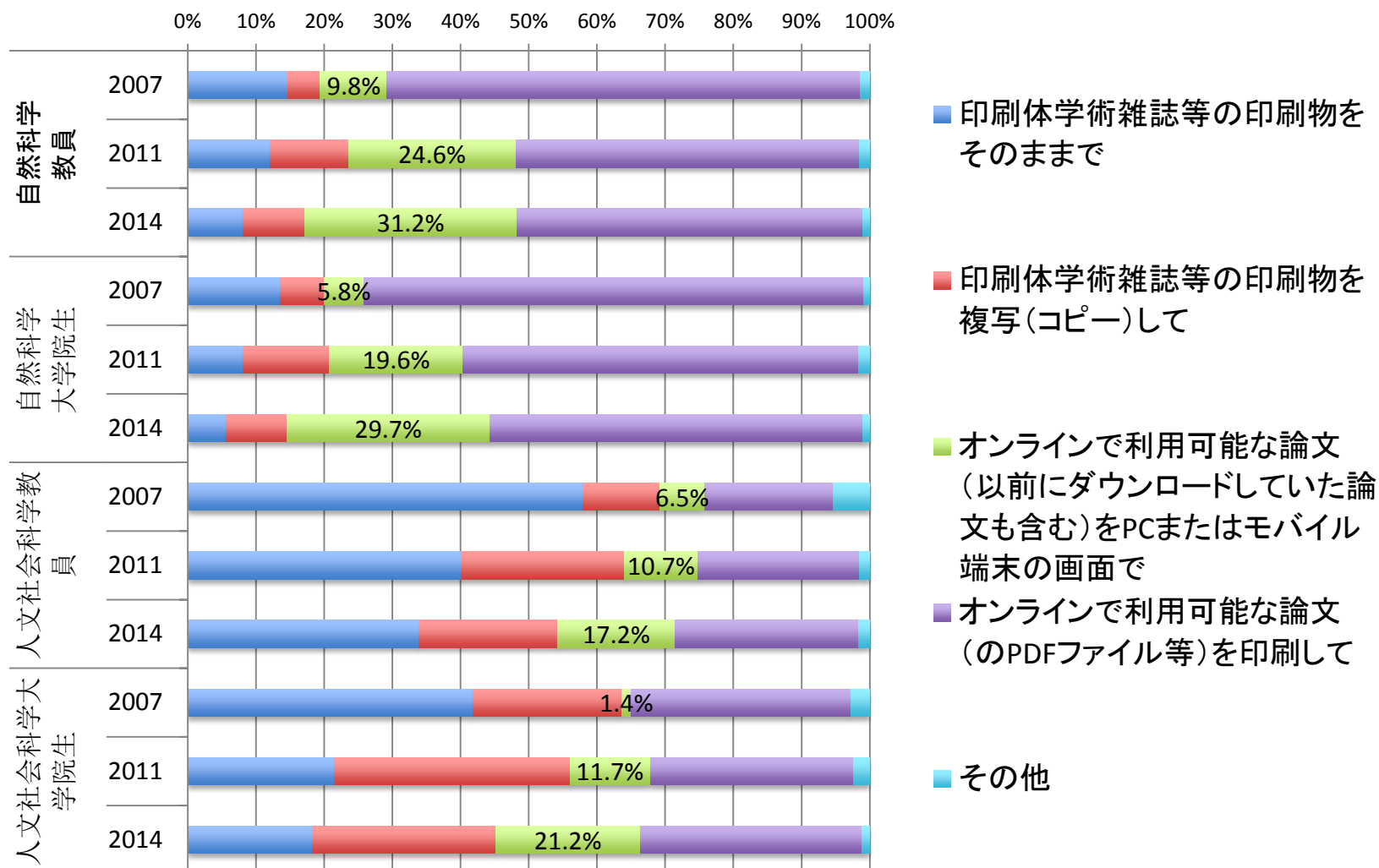
自然科学



人文社会科学



最後に読んだ論文の読み方 (2007 - 2014)



図書館外での進行

- 出版社によるリサーチワークフローへの関心
＝ 図書館市場の迂回
 - － DRM技術を利用したシェアリング環境の提供 (Wiley, NPG等 --- ReadCube)
 - － MendeleyやSSRNの買収, BioRNの開始 (Elsevier)
 - － CHORUS (ClearingHouse for the Open Research of the United States”)
- 負の遺産としてのSci-Hub問題
 - － & Library Genesis (LibGen)
 - － 認証という「信頼の枠組み」の問題も提起

電子書籍

- 高い期待, しかし実態が伴わず(構造的問題)
 - 個人向け市場は, Amazon kindleの一人勝ち
- しかし
 - 「ファウスト的取引」(Clifford Lynch)
 - ライセンシング(契約期間, ファーストセール・ドクトリンの否定), DRM, 機能の制限
 - 学術図書館市場と個人市場の分離
 - 学術図書館では, 買い取り(データローディング)またはリモートアクセス
- 利用可能タイトル数は増加, しかし米国での提供タイトル数は学部生用図書館の平均で約8万, 大学院・教員用図書館の平均で約14万程度(FY2012)

学術論文へのオープンアクセス

- 米国2014年統合予算法修正案
 - 米国教育省、労働省、健康・人的サービス省に属し、年間の研究予算が1億ドルを超える機関に対し、パブリックアクセス方針を定めるよう要求(エンバーゴ期間: 12カ月)
- 欧州委員会 "Horizon 2020"
 - 同プログラム下の助成研究の成果に関連した査読論文は、1) 6ヶ月以内(社会科学, 人文学では12ヶ月以内)にセルフ・アーカイビング, 2) OAジャーナルまたはハイブリッドジャーナルに掲載, のいずれかの方法で公開することを要求
- フィンチレポート
 - ゴールドOA(ハイブリッドとpure OAジャーナルの両方)の重視 – 批判と影響
 - JISC APC
- SCOAP³
 - 図書館が従来「購読料」として支払っていたものを対象雑誌の「出版料」に振替えることで、世界中の誰もが無料でそれらの学術雑誌の論文を読むことができるオープンアクセス化の実現を指向

オープンアクセス出版： 理念からビジネスへ

- 意義
 - 「市場の二面性」の打破，すなわち著者サイドの市場で課金が行なわれることで出版者間の競争を期待（？）
- Full (Pure) OAジャーナル vs Hybrid OAジャーナル
 - Full OAジャーナル専門出版社 + 学会系等
 - 大規模商業出版社の対応

オープンアクセス出版： 理念からビジネスへ

- メガジャーナル(?)
 - 「非選択的」、「インパクト・ニュートラル」、「厳密だが包括的な査読」；規模は大きくないが、同様の編集方針を持つ雑誌群の存在 (Binfield, 2013)
- 引き続く新たなOAジャーナルの創刊
 - Science Advances (AAAS), Royal Society Open Science (The Royal Society), Cogent Economics & Finance (Cogent OA – Taylor & Francis), IEEE Access, ...
 - 学会系雑誌の商業出版社プラットフォームへの移動(OAジャーナルとして) 例. [Wiley OA New Titles](#)
- ピアレビューの変化
 - カスケード査読
- フリップング・ジャーナル
 - オフセット(相殺)モデル
 - “Gold for Gold”プログラム (Royal Society of Chemistry)
 - IoP Publishing in Australia

Peer review cascade



Genome **Biology**

High
rejection rate



BMC
Bioinformatics



BMC
Evolutionary Biology



BMC
Genomics

Moderate
rejection rate



BMC
Research Notes

Low
rejection rate



BioMed Central
The Open Access Publisher

Matthew Cockerill, “10 years of Open Access at BioMed Central.”

<http://www.slideshare.net/BioMedCentral/10-years-of-open-access-at-biomed-central>

オフセットモデルの進展

- 「購読契約 with APC支払い」
- VSNU(オランダ大学協会) with Springer, Wiley, and Elsevier, 英国
- Max Planck Instituteの提案, Berlin 12

課題と展望

- APCの存在に関わりなく、長期にわたって雑誌を継続的に出版するための適切な財源の確保が必要
- APCによる資金調達モデルの利点の一つは、投稿数と出版論文数が増大した場合に、うまく調整できること(他の資金調達モデルは、投稿数の増大に対処できない可能性が高い)
- オフセットモデルは、出版社が購読機関との関係を保持しながら徐々にOA出版モデル(APCビッグ・ディール)に移行する道である可能性
- オフセット契約は出版社にとって容易な販売方式であり、価格が購読のみの契約と同じままであるなら購読者の利益に(最も利益を受けるのは民間企業)
- しかし、オフセットモデルの契約拡大は、中小規模のOA出版社や学会に対し著者獲得の競争で大きな不利益をもたらす可能性が高い(さらなる寡占化のリスク)

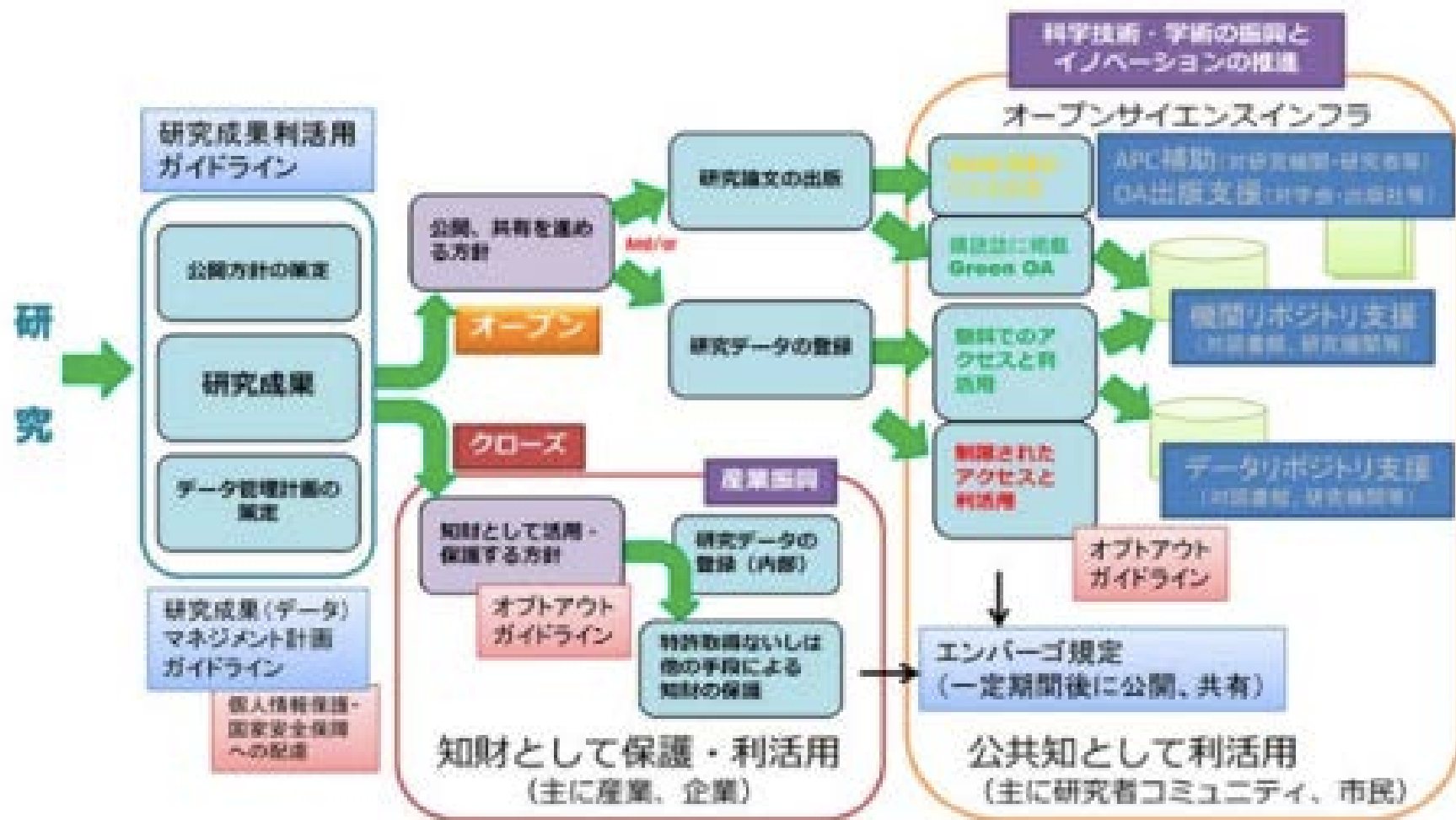
※OA出版やオフセット契約が出版者間の競争を促進するのか、あるいは独占を推進するのか

我が国におけるオープンサイエンス 推進のあり方について

～サイエンスの新たな飛躍の時代の幕開け～

2015年3月30日

[内閣府] 国際的動向を踏まえたオープンサイ
エンスに関する検討会



下記図表を参考に和訳、改定

Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020 Version 1.0 11 December 2013 p.4

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

国際的動向を踏まえたオープンサイエンスに関する検討会『我が国におけるオープンサイエンス推進のあり方について』内閣府, 2015.3.30

<http://www8.cao.go.jp/cstp/sonota/openscience/index.html>

オープンサイエンス

- 「オープンアクセスとオープンデータを含む概念」
(p. 3)
- 対象：「公的研究資金による研究成果として得られた論文や研究データ(p. 15)」
- 公的研究資金
 - 1) 競争的研究資金及び公募型の研究資金
 - 2) 国費が投入されている独立行政法人及び国立大学法人等の運営費交付金等

論文	原則公開
論文のエビデンスとしての研究データ	
その他研究成果としての研究データ	可能な範囲で公開

オープンサイエンス推進における 大学図書館の役割

- 図書館・機関リポジトリがオープンサイエンスの基盤(p. 14, p. 19の図)
- 「大学図書館職員等を中心としたデータ管理体制」(p. 21)
- 「論文や研究データの永続的, 長期的保存を担保するために...大学図書館... 等の協力を得ることが有効」(p. 21)

国際的な取り組み

- G8 オープンデータ憲章 (G8 Science Ministers Statement [on Open Scientific Research Data]. 2013.6)
 - 原則1 : 原則としてのオープンデータ
 - 原則2 : 質と量
 - 原則3 : 全ての者が利用できる
 - 原則4 : 改善されたガバナンスのためのデータの公表
 - 原則5 : 技術革新のためのデータの公表

<http://www.kantei.go.jp/jp/singi/it2/densi/dai4/sankou8.pdf>
- RDA (Research Data Alliance)
 - 2013年3月発足
 - “Data sharing without barriers”

研究データ管理とオープンアクセス

- 研究データ共有
 - 「データ共有とは、他者が利用できるよう研究データを公開すること」(Borgman, 2012)
- 背景：
 - デジタル技術による、データの保存, 共有, 再利用の可能性の向上
 - データの量的増大

研究データの共有を後押しする原理

- 公的資金を受けた研究成果の還元
- 研究の再現または検証(e.g.「STAP細胞」問題)
- 既存のデータを用いた他者による新たな課題の研究(メタ分析, マッシュアップ, 大量データのマイニング等)
- 研究と革新の進展:「第四のパラダイム」としての計算科学による一連の新たな方法等

Borgman, Christine L. “The conundrum of sharing research data,” *Journal of the American Society for Information Science and Technology*, 2012, vol.63, no.6, p.1059-1078.

研究データへのオープンアクセスの始まり

- World Data Center (WDC) system (1958 –)
 - 1957～1958年の国際地球物理学年の観測プログラムで収集されたデータのアーカイブと提供のために、ICSU(国際科学会議)によって設立
- CODATA (Committee on Data for Science and Technology; 科学技術データ委員会)(1966 -)
 - ICSU(国際科学会議)によって設立
 - 基礎定数についてのCODATAタスクグループ(1969 -)
 - 国際的に認められた基礎物理定数および関連した変換係数の組を定期的に提供

国際的な研究データ共有の例

- ヒトゲノム計画 (Human Genome Project)
 - 人ゲノムの20,000以上の遺伝子配列を特定し染色体上にマッピングし、そしてDNAを構成する化学塩基の組み合わせ配列を明らかにする国際的取り組み。1990年に開始され、計画は2003年に完了。世界中のさまざまな場所のDNA配列を保存する分散型データベースであるGenBankの構築
- タンパク質構造データバンク (Protein Data Bank), GEON (地球科学), 天文学における総観天空探査 (Sloan Digital Sky Survey等) ほか

研究データ管理計画の要求

- 2011.2 NSF すべての研究資金申請における「データ管理計画(Data management plan)」の提出要求
 - あくまでデータ管理計画であってデータ共有計画ではないが、共有を強く求めるとともに計画もピアレビューの対象とされているため強い影響力
 - 2003.2 NIH 研究データの共有に関する声明(単年度の補助金申請額が50 万ドルを超える申請者に契約担当者との交渉を要求)
 - NSF においてもNIH よりも早い時期から資金提供契約でデータ共有が奨励されてきたが、強制力を伴わなかった
- 2011.6 全米人文科学基金(National Endowment for Humanities)のデジタル人文学部門 NSFと同様の要求を発表
- 2013.2 米国大統領府科学技術政策局(Office of Science and Technology Policy, Executive Office of the President) 研究開発支出額が年間100 万ドルを超す政府機関に対し6 ヶ月以内に連邦予算による研究成果(ピアレビュー出版物と機密研究以外の研究データ)へのパブリックアクセスを拡大するための計画案を提出するよう命令
- 2013.5 政府情報のオープンデータ化を義務付ける大統領令

データキュレーション

「科学，社会科学，人文学全般の研究教育活動に関係を持つ有益なデータのライフサイクル全般にわたる積極的で継続的なマネージメントである。データキュレーション活動は長期に渡って，データの発見，検索，品質維持，価値の向上再利用のための提供を可能にする。この新たな領域には，説明，アーカイビング，認証，管理，保存，検索，利用が含まれる。」

DCEP: Digital Curation Education Program.

<http://cirss.lis.illinois.edu/CollMeta/dcep.html>

英国におけるデータポリシー

- RCUK「データポリシーに関する共通原則(Common Principles on Research Data Policy)」: できる限り制約なしに適時のかつ責任ある方法でデータの公開を強く要求

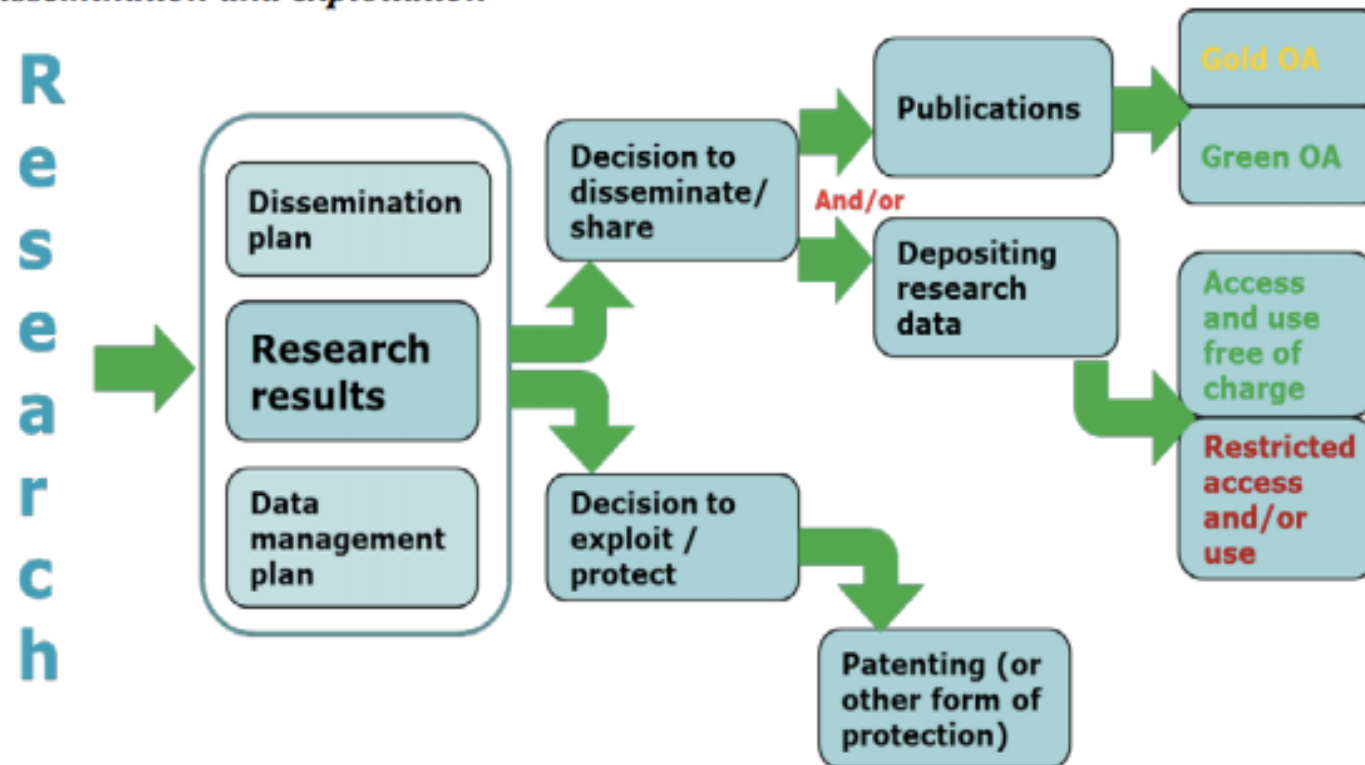
● Full Coverage ◐ Partial Coverage ○ No Coverage

Research Funders	Policy Coverage		Policy Stipulations					Support Provided			
	Published outputs	Data	Time limits	Data plan	Sharing/ access	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	◐
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
EPSRC	●	●	●	◐	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	◐
Cancer Research	●	●	●	●	●	●	●	◐	●	○	●
European Commission	●	●	◐	●	◐	◐	◐	●	●	◐	●
Wellcome Trust	●	●	●	●	●	●	●	●	●	●	●

英国の資金提供団体の論文と研究データの公開方針. Digital Curation Centre. “Overview of funders' data policies.” <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>

欧州委員会 “Horizon 2020” におけるオープンアクセス

Graph: Open access to scientific publication and research data in the wider context of dissemination and exploitation



より広い範囲の流通および展開の文脈における科学出版物と研究データへのオープンアクセス. European Commission. *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*. Ver. 1.0 2013.12, 14 p.

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-

学術雑誌における研究データへの注目

- データジャーナルの創刊
 - Scientific Data (NPG; 2014.5 -), GigaScience (BioMed Central; 2012 -), Biodiversity Data Journal (Pensoft, 2013-), Geoscience Data journal (Wiley; 2014 -) 等 cf. MethodsX (Elsevier; 2014 -)
- Science: データに関する特集号 (2011.2 special issue “Dealing with Data”)
 - 「データの形成または分析に関連する」コンピュータプログラムの共有; 論文中での「データの利用可能性やキュレーションに関する特別な説明」
- 「進化と生態学の主要雑誌」新たなデータ・アーカイビングの方針 – 公開アーカイブへのデータのデポジットを要求または奨励; The American Naturalist, Evolution, Journal of Evolutionary Biology, Molecular Ecology, and Heredity
- 経済学その他でもデータへのアクセスを要求
 - 方式 フォーマル (Protein Data Bankへの登録へのデポジット, 論文には構造エントリ番号を記載), インフォーマル (資源へのリンク)

学術雑誌における研究データへの注目

- PLOS: 論文の一部としてData Availability Statementを要求
 - (データの公開方法(リポジトリでの公開[強く推奨], 関連情報ファイルの添付, 要請に応じて提供, サードパーティから提供)についての説明;なお, リポジトリでの公開の場合はDOIやアクセション番号の付加)を要求(2014.3 -)
- “stub article” (stub; 切り株?; wikipediaでは展開が必要な短い記述のこと)
 - 「分野別リポジトリでの重要かつ新たなデータセットの寄託を公表し, その(データ)生成に関わる背景情報を提供はするが, データの分析を行わず後続の出版物に委ねる(論文)」(Lynch, Clifford A. “The shape of the scientific article in the developing cyberinfrastructure,” *CT Watch Quarterly*. 2007.8

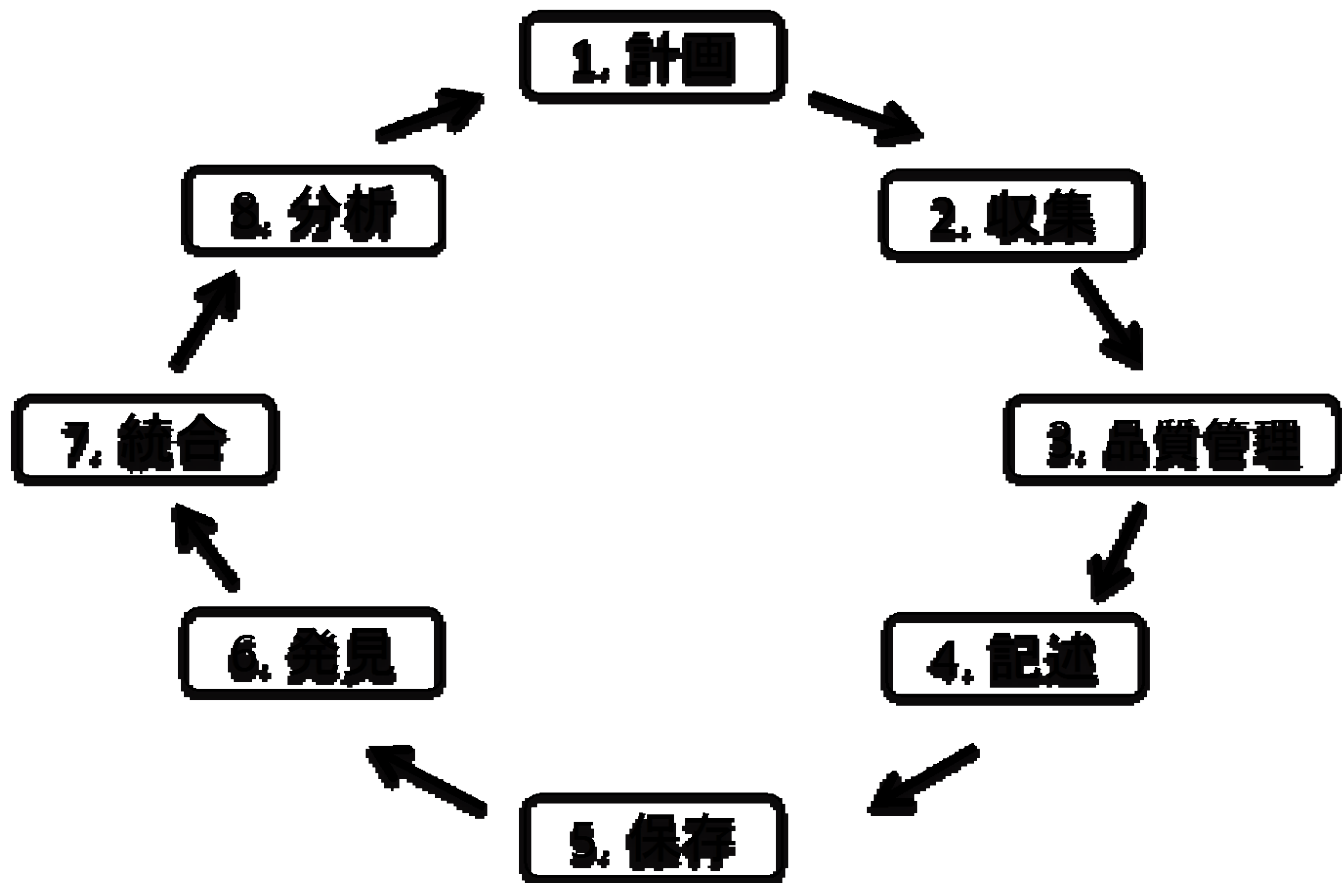
研究データ管理サービス (Research Data Management Services)

- 研究データ管理サービスとはデータのライフサイクル全体を対象として、「図書館がデータ管理に関連して研究者に提供するサービス, 情報のサービス, および技術サービスを含む」
 - サービスの例.
 - データ管理計画あるいはメタデータ基準についての教員, スタッフ, あるいは学生との相談
 - 知見やデータセット引用へのレファレンス支援
 - データやデータセットのためのウェブ上の案内や発見支援
 - データリポジトリのための技術支援
 - リポジトリのためのデータセットの準備
 - データセットのリポジトリからの登録解除または除外
 - データセット用のメタデータの生成

Tenopir, C.; Birch, B.; Allard, S. *Academic Libraries and Research Data Services: Current Practices and Plans for the Future. An ACRL White Paper*. Association of College & Research Libraries. 2012, 54p.

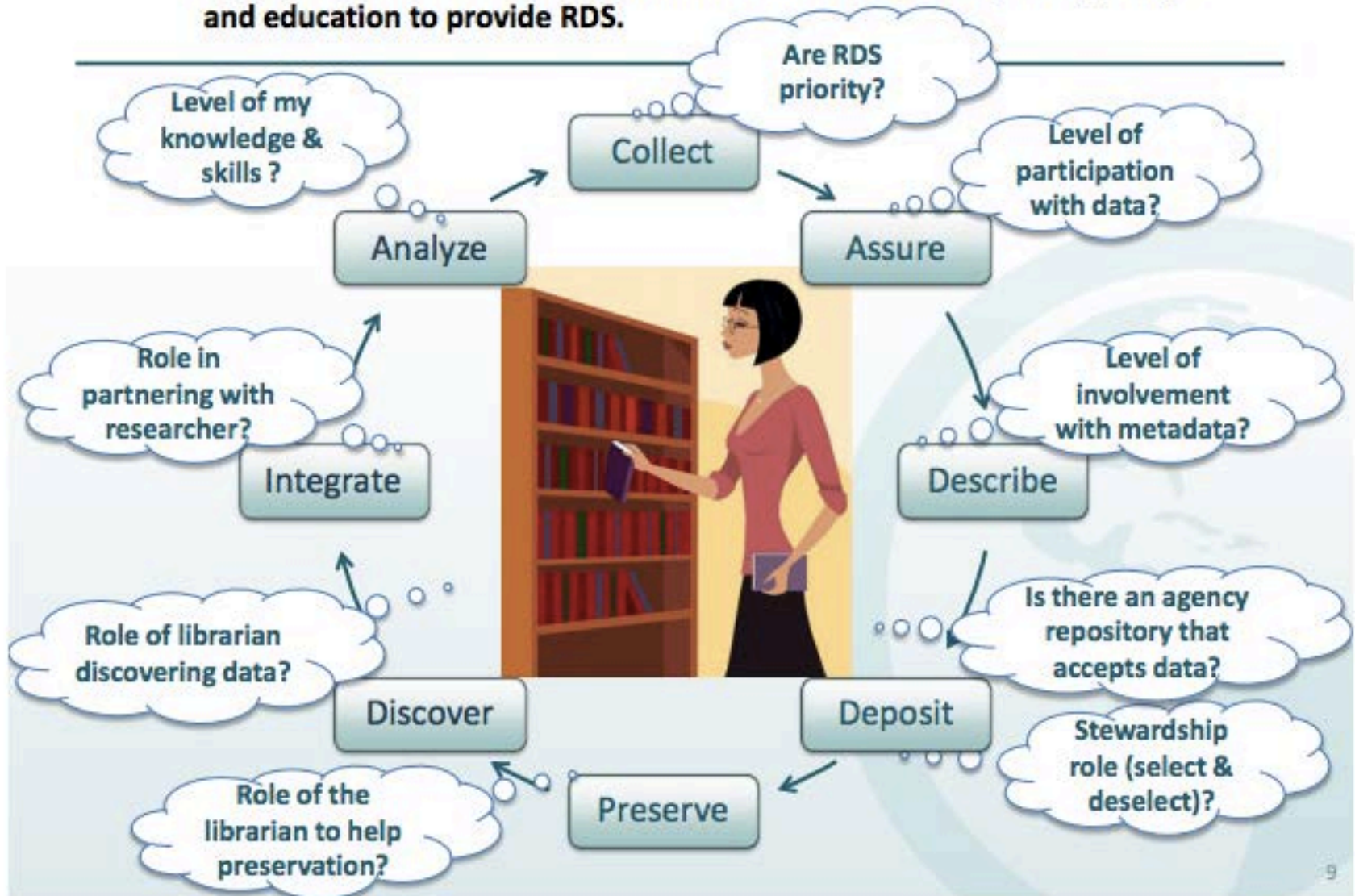
<http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitenpapers/Tenopir>

データのライフサイクル



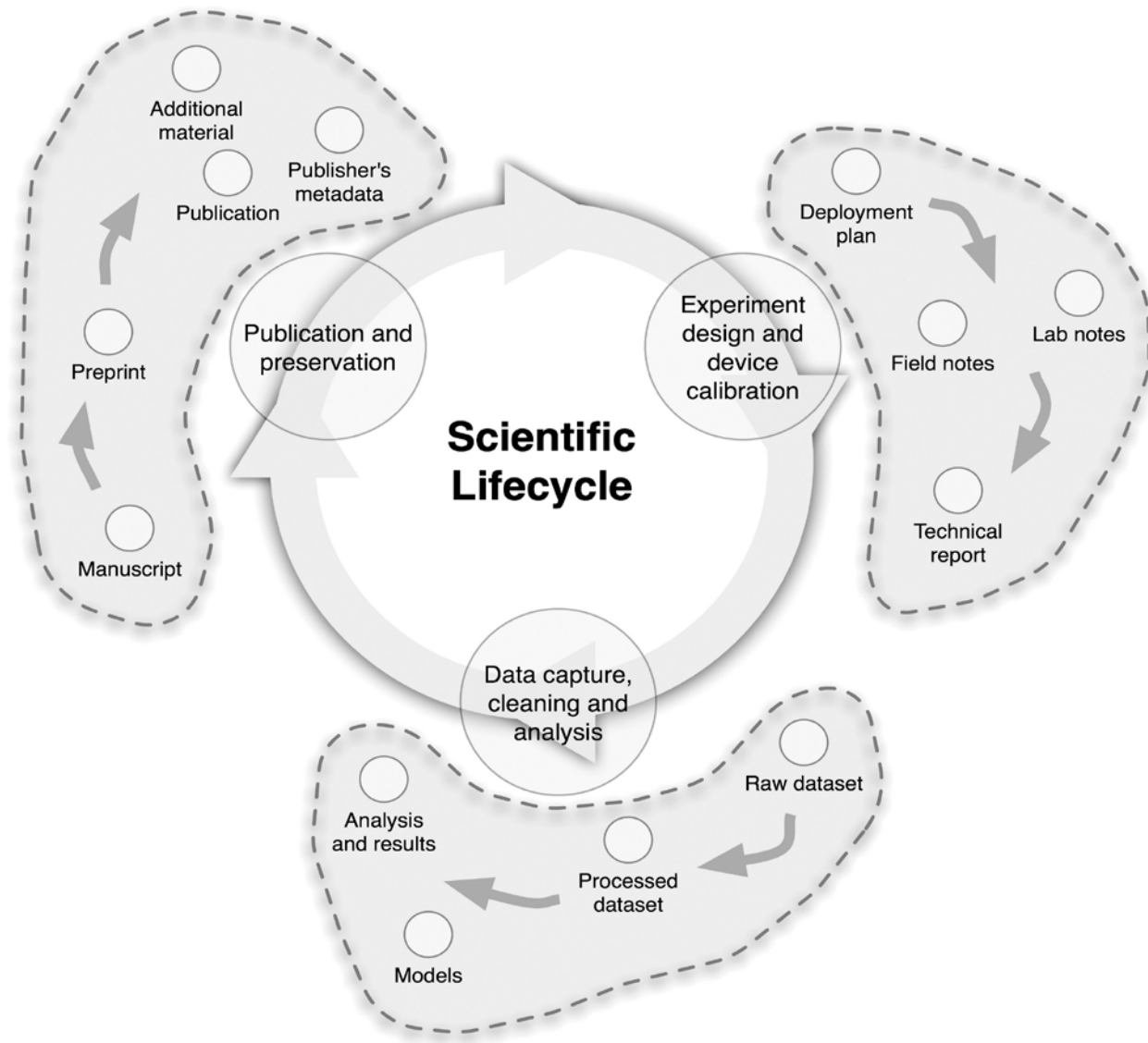
(<http://www.dataone.org/best-practices> をもとに作成)

Figure 2. The librarian ponders whether she has the background, skills, and education to provide RDS.



Tenopir, C.; Birch, B.; Allard, S. *Academic Libraries and Research Data Services: Current Practices and Plans for the Future. An ACRL White Paper*. Association of College & Research Libraries. 2012, 54p.

<http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Teno>

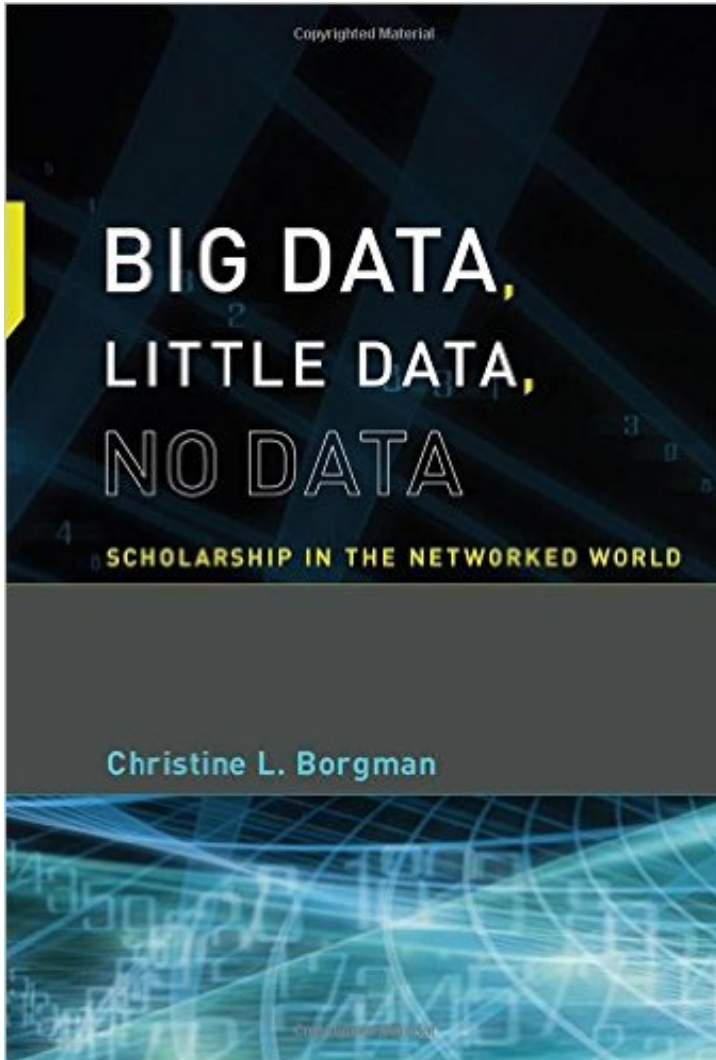


Pepe, A., Mayernik, M., Borgman, C. L., Sompel, H. V. From Artifacts to Aggregations: Modeling Scientific Life Cycles on the Semantic Web. *Journal of the Association for Information Science & Technology*, 2010, vol. 61, issue 3, pages 567-582

データの公開と共有の本質的課題

1. データとは何かが不明確
2. 学術論文の場合とは異なる権利関係(著作権, プライバシー, 所有権, 特許権, 法的責任, ...)
3. データ利用は文脈依存, メタデータと説明が不可欠, しかし作成には時間を要する; 誤用, 誤解釈, 法的責任 --- 「データは資産でも債務でもある」
4. 「データ出版」という単純化の問題
5. データは一つ一つ異なる(出版物との違い)
6. 保存とアクセスのバランス

データ公開，共有の課題



Christine L. Borgman.
Big Data, Little Data, No Data: Scholarship in the Networked World. MIT Press, 2015.1

データ公開，共有の課題

1) データとは何かはあいまい

－ 多様なデータ，データセット

- 社会科学のサーベイ調査や，人文学のテキスト・コーパスを含む

－ NSFは研究データの定義を回避

- データ管理計画において，説明をFAQの最初に置くことでデータの定義を回避

「Q. データ管理計画の対象となるデータの構成要素は何か？」

「A. データを構成する要素は，ピアレビューとプログラム管理の過程を通じて関係コミュニティによって決定されることになる。データ，出版物，サンプル，物的コレクション，ソフトウェア，モデルが含まれ得るが，これらに限定されるわけではない」

－ したがって，何を保存するか，公開するかもあいまい

データとは何か

- 人工物, または観察記録, Michael Bucklandの簡潔な表現を用いるとせいぜい「不確かな証拠」(Buckland, 1991)
- 見る人の目にのみ存在するのかもしれない; 観察記録, 人工物, あるいは記録がデータを構成することを認めること自体が学術活動である
- 最も広く引用される定義 (National Academies of Science Report, p.15)
 - 「データは, 対象物, アイデア, 状態, 状況, または他の要素を記述する事実, 数字, 文字列, シンボルである」
- より最新の実用的定義 (国際学会の資料) – [共有の議論に有用]

本文書で用いる「データ」という用語は, 広く包括的であることを意図している。デジタルによる文献の表現(テキスト, 音声, 静止画, 動画, モデル, ゲーム, あるいはシミュレーションを含む)に加えて, 分光学, ゲノム配列, 電子顕微鏡法のデータといったさまざまなタイプの実験室データ, リモートセンシング, 地理空間, 社会経済のデータといった観測データ, そして人間または機械によって生成または編集された他の形式のデータといった, 一般に有用とするためにコンピュータの機器やソフトウェアの活用を必要とするデータやデータベースの形式もあてはまる。(Uhlir & Cohen, 2011)

Open Dataの定義

- 1) だれでもが自由に利用, 再利用, 再配布できるなら, そのデータはオープンである。最大でも専門分野に限られ, 帰属表示および／または継承の要件がある。

Open Data Commons. (2013). <http://opendatacommons.org/>

- 2) 知的なオープン性の基準に見合うデータ。データは, 入手可能, 利用可能, 評価可能, 理解可能でなければならない。

Boulton, G., et al. (2012). "Science as an open enterprise." The Royal Society. <http://royalsociety.org/policy/projects/science-public-enterprise/report/>

- 3) オープンであること, 柔軟性, 透明性, 法律遵守, 知的所有権保護, 公的責任, 専門家意識, 相互運用性, 品質, セキュリティ, 効率, 説明責任, 持続可能性。

Organization for Economic Cooperation and Development. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. <http://www.oecd.org/dataoecd/9/61/38500813.pdf>

データ公開，共有の課題

2) 学術論文との違い

✓ 権利関係

- 著作権
- その他(プライバシー, 所有権, 特許権, 法的責任, ...)

✓ 公開のインセンティブ

※雑誌論文は読者への供給のためにパッケージ化されているが、データは学術活動のプロセスから解放するのが難しい。データの公開には、多くの場合、研究実施と公表物の執筆に対するかなりの投資が求められる。データは、研究生活の過程全般を通して蓄積される貴重な資産であり、仮に公開する場合でも注意深く行われるべきである

学術論文のオープンアクセス

- 「オープンアクセス文献は、デジタルでオンラインの、無料の、そしてたいていの著作権やライセンスの制限から自由なものである」
- 含意されていること
 - a. 出版者といった別の団体へ所有権を移転しない限りおよび移転するまでは、著者が自らの著作の著作権者であること
 - b. 研究者は基本的に研究論文の執筆に対し報酬を受けることはない。研究者は自らの論文を収益を失うことなく広く頒布することができる

※Green OAもGold OAも、上記の原則に依拠

※研究データの場合は、両方ともあてはまらない

By "open access" to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.

Budapest Open Access Initiative (2002)

<http://www.budapestopenaccessinitiative.org/read>

データ公開，共有の課題

3) データの利用は文脈依存

- ✓ メタデータと説明が不可欠
 - しかし，作成には時間を要する
 - 誰に向けて説明？（説明を理解できるレベル）
- ✓ データの利用には特別な機器やソフトウェアなどが必要になる場合も
- ✓ 誤用，誤解釈，法的責任 --- 「データは資産でも債務でもある」
- 「“実在しない場所からの視点”などない。知識は常に，場所，時間，状況，慣習，理解に根付いている。単一の知識ではなく，複数の知識が存在する」

Van House, N.A. (2004). “Science and technology studies and information studies.” in Cronin, B. (Ed.). *Annual Review of Information Science and Technology*. Medford, NJ, Information Today, 38, 3–86.

データ公開，共有の課題

4) データ出版という単純化の問題

- 学術出版の三つの機能：
 - i. 正当化 (legitimization)；ピアレビューを通じて
 - ii. 頒布 (dissemination)
 - iii. アクセス，保存，キュレーション
- データは一つ一つ異なる(出版物との違い)
- データ自体が査読を受けているわけではない
- データ出版 (data publication) というメタファーは，国際統計，国勢調査等の狭い意味でしか成立しない
- 雑誌論文とデータセットの間の1対1のマッピング
 - 発見を容易にはするが，関係はしばしば多対多

データ公開，共有の課題

5) 保存とアクセスのバランス

- 保存に最も有効な形式は，アクセスには向かない(i.e. 高解像度 vs 低解像度)
- デジタル保存： ダークアーカイブ等，研究コミュニティ，大学コンソーシアム，あるいは国全体でのインフラ投資が不可欠

6) 保存のコストは学術論文の場合よりも高くつく

データ公開，共有の課題

「ビッグデータ，リトルデータ，そしてノーデータの概念さえも，きちんと理解されないままである。より良いデータ管理，共有，認定，帰属を促進する取り組みはかなり志向されているが，利害関係者はその出発点，最終目標，そしてそこに至る道筋に賛成していない。データと呼ばれるものは何かについての合意が欠けており，すべての領域にまたがるデータ・スカラーシップの多様性に対応する共有，公開，デポジット，認定，帰属表示，引用，アクセス維持の政策を確立するのは困難なままである。データへのアクセス維持は困難で費用のかかる作業であり，コストと便益は利害関係者間に偏って分布する。何を保持するか，そしてその理由に関する疑問は，誰が，どのように，なぜ，誰に対して，どれくらいの期間，データを保持していくかという疑問と分離できない。個々の研究者，学生，図書館員，文書館員，研究担当者，あるいは雑誌編集者は，この巨象の難問に対してせいぜい蟻の目を持つに過ぎないのかもしれない。」

Christine L. Borgman. *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press, 2015.1

図書館および機関リポジトリとの関係

- 1) なぜ研究データを扱うのか (例. University Record)
- 2) 「データを確実にアーカイブすることがオープンアクセスへの第一歩」 (例. DataVerse, SciDrive; 保存は行っても、可能な場合にだけ公開できる)
- 3) 知識インフラの構築
 - 知識インフラの特性
 - 障害が起きるまで、存在に気づきにくい
 - コストや努力が見えにくい
 - 「集合的な」取り組みの必要性
 - 識別子管理 (DOI, ARK, PURL等)

知識コモンズの経済学

		減少可能性(競争性)	
		低	高
排除可能性	困難	<u>公共財</u> オープンアクセス・リポジトリ (有用な知識, タ焼け空)	<u>共有資源</u> 図書館(紙媒体) (会員制のデジタル共有資源サイト, 灌漑システム)
	容易	<u>自然独占</u> 電子ジャーナル・プラットフォーム (電力会社, 鉄道会社)	<u>私的財</u> 個人の蔵書(紙媒体) (パソコン, 衣服, アイスクリーム)

Borgman, Christine L. *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press, 2015.1, 383 p.; C. Hess & E. Ostrom (Eds.), *Understanding Knowledge as a Commons: From Theory to Practice*. MIT Press, 2007, p. 9.; N・グレゴリー・マンキュー『マンキュー経済学. 第2版①ミクロ編』東洋経済, 2005, p. 301 – 321.

「コモンズの悲劇」

- Hardin, Garrett. “Tragedy of commons,” *Science*. Vol. 162: 1243-1248 (1968)
- 人々が限られた資源を共有し、そして、人々が合理的な選択をするならば、共有資源の持続可能性が失われてしまう
 - － 例. 牧草地, 魚の乱獲

アンチコモنزの悲劇

- Heller & Eisenberg, “Can Patents Deter Innovation? The Anticommons in Biomedical Research” *Science*. Vol.280, 698 (1998)
- 知識の私有化(Privatization)は、コモنزの悲劇を解決はしたが、アンチコモنزの悲劇という新たな悲劇を生み出したと指摘
- 共有されるべき財産が細分化されて私有され、社会にとって有用な資源の活用が妨げられる

デジタル化による 公共財の性質の変化

- 「この以前には捕捉できなかったものを捕捉する能力は、資源を非競合的かつ非排他的な公共財から、持続可能性と保存を確実にするために管理、監視、保護が必要な共有資源へと(以前と比べて遥かに容易に)移動させるという、資源の性質の根本的な変化をもたらす。」

学術コミュニケーションの変化：まとめ

- 学術論文のデジタル化（電子ジャーナル）
 - ビッグ・ディール
 - 利用者と図書館との接点の希薄化
 - オープンアクセス
 - 機関リポジトリにおける一定の成果
 - OAジャーナルの進展
 - 今後、大学図書館はどのように関与？
 - 研究プロセス自体の変化
 - 「学術的記録」の範囲の広がりと量的拡大
 - 「学術的記録」の利用や保存を誰が担うのか、支援は必要なのか、その場合どのような内容なのか？
- ⇒ “学術情報システム”の再考の必要性