

レポート

欧州オープンサイエンスクラウドに見る オープンサイエンス及び研究データ基盤政策の展望

科学技術予測センター 客員研究官・情報通信研究機構 村山 泰啓
科学技術予測センター 上席研究官 林 和弘

概要

オープンサイエンス、研究データ共有（シェアリング）に関する議論が、政策面からも、技術・科学の推進面からも国際的に議論されている。第5期科学技術基本計画では、内閣府の検討会の議論も踏まえて「オープンサイエンスの推進」の項を設定している。その実現に際しては多種多様な研究データの利活用が重要となる。現在、国際的な政策上の論点の中でもデータの相互利用、相互運用性（data interoperability）やデータ・パブリケーションやデータの保存機関・リポジトリに加えて、データ利用環境となる研究データ基盤（Research Data Infrastructure）に注目が集まっている。欧州にて注目を集めている「欧州オープンサイエンスクラウド」計画を中心にしつつ、他の研究データ基盤整備やデータ相互利用体制にも着目しながら紹介し、今後の取組のポイントを論じる。

キーワード：オープンサイエンス、研究データ共有、オープンデータ、データリポジトリ、研究データ基盤、RDA, OECD, G7

1. はじめに

2013年のG8首脳会合、科学技術大臣会合において研究データのオープン化が合意されて以来、オープンサイエンス、研究データ共有（シェアリング）に関する議論が、政策面からも、技術・科学の推進面からも国際的に議論されている。我が国でも、内閣府の「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」において、特に科学技術に関する研究データの共有やオープン化について焦点を当てて議論が行われ、報告書が2015年3月にまとめられた¹⁾。

また第5期科学技術基本計画（2016年1月閣議決定）では、内閣府の検討会の議論も踏まえて第4章を「科学技術イノベーションの基盤的な力の強化」にあて、その中に「オープンサイエンスの推進」の項を設定している。本文中では“オープンサイエンスの推進体制を構築し、公的資金による研究成果については、その利活用を可能な限り拡大することを、我が国のオープンサイエンス推進の基本姿勢”と述べている。

オープンサイエンスの実現に際しては多種多様な研究データの利活用が重要となるが、その利活用におい

て、現在、国際的な政策上の論点の中でもデータの相互利用、相互運用性（data interoperability）やデータ・パブリケーション²⁾やデータの保存機関・リポジトリの在り方など³⁾に加えて、データ利用環境となる研究データインフラストラクチャ、研究データ基盤（Research Data Infrastructure）に注目が集まっている⁴⁾。

本稿では、近年重要視されているオープンサイエンス政策の国際動向について、欧州にて注目を集めている「欧州オープンサイエンスクラウド」計画を中心に、他の研究データ基盤整備やデータ相互利用体制にも着目しながら紹介し、今後の取組のポイントを論じる。

2. オープンサイエンス政策の推進と課題

2-1 オープンサイエンスの推進と課題

我が国でオープンサイエンス、特に現在その中心的な概念であるデータ共有について議論していく上で、これまでに幾つか課題が挙げられている。例えば、研究論文・ジャーナルへのオープンアクセスの問題と同一視されること、「オープン化」という言葉から無制限な

データの公開と誤解されること、同様に知的財産権などの関連で企業活動や個人情報などと無縁と誤解されること、などが挙げられるだろう。これは欧米でもまだ整理されていない問題で、同様な混乱が欧州内の会議でもあると聞く。つまり、日本国内と欧州で同時に一体的に先端的な議論が進行中であるとも言える。

サイエンス（ここでは人文・社会科学全てを含める、もっとも広い意味での科学又は学問）は、研究プロセスや結果を学会・学術誌等で発表して（出席者・読者と「共有」して）初めて健全な相互批評や評価、活用の対象となる。現代の研究の多くがデータに基づいていることから、こうした健全な相互批評を経たコンセンサス形成について、データの共有が重要という指摘は少なくない⁵⁾。研究データの共有化は今後も何らかの形で考慮していく必要があるだろう。一方、例えば企業活動や個人のデータ、実施中の研究で成果が出る前のデータ、など、データ公開に配慮が求められるケースもある。また、成果に結び付かなかったデータなどオープン化のインセンティブが働きにくいケースもある。現場の負担を過度に増やさず、長い目で見てそれぞれのコミュニティ・業界の活動の減速を防ぐ、又は加速するデータ共有の方針が重要である。

2-2 デジタルデータの相互利用（データ・インターオペラビリティ）環境の整備

オープンサイエンスの枠組みでデータが有効に活用される社会においては、データ公開とは別に、データが相互運用可能であること（インターオペラビリティ；interoperability）や標準化（standardization）が極めて重要と考えられている。現在も、文書データや画像データは限られた形式で流通されるため、多くの場合、自動で判別・読み取り・可視化が可能であるが、研究データはこれと比べれば想像を超えて多様であり、現状ではかなりの部分を人手によらなければ処理が実現しないものが多数ある。将来の情報社会において、識別子や分野横断型のメタデータなどを使ってデータが自由に検索、追跡可能であり、またデータの信頼度（trustworthiness）、データ形式や分野固有の処理方法等を見付けるための世界統一レジストリが整備されているとすれば、ユーザの指示を受けたソフトウェアは、各種データが一定の信頼度があると判断した上でその所在、処理方法を自ら発見して必要な処理モジュールを判断して処理を行える（マシン・アクションナブル）可能性がある。

このような環境やデータの整備をすることで、データを自分だけで専有して処理する場合も、他人と共同で解析する場合も、また一定の経緯を経て一般公開する場合も、ユーザはいつもほぼ同じように処理が可能

であり、発見・読み込み・処理のための準備は非常に容易となる。研究データは「ロングテール」⁶⁾と例えられるように非常に多様で、異種少量のデータセットが数多くある。このように幅広いデータについて、将来飛躍的に管理・利用を向上させると期待されるような、識別子、メタデータ、国際レジストリ、使用用語（vocabulary）の統一整備などを進めるための調査・検討が、G8 国政府が関与してきた国際組織 RDA（Research Data Alliance）などで進められている。こうした検討の場には、かつて TCP/IP 開発の場にいた国際的に著名な有識者も参加している。ICT の検討対象が時代とともに変化していることを示唆する。

このような概念検討、デザインを徐々にまとめつつ、一方では実際に、これを実践すべき科学データインフラストラクチャを整備し、新たなビッグデータ解析や科学的発見、経済効果などの具体化を目指す動きが加速しつつある。幅広いデジタル情報が活用される現代では、こうしたインフラ整備と活用が社会・経済にまで影響を及ぼすと考えられている。次節中で述べる欧州政策の議論のように、科学界が情報の利活用を進めるスキルの開発と成功実践例を生み出すことで、より広いセクターでの活用をリードする「アーリーアダプター」となることが期待されている。

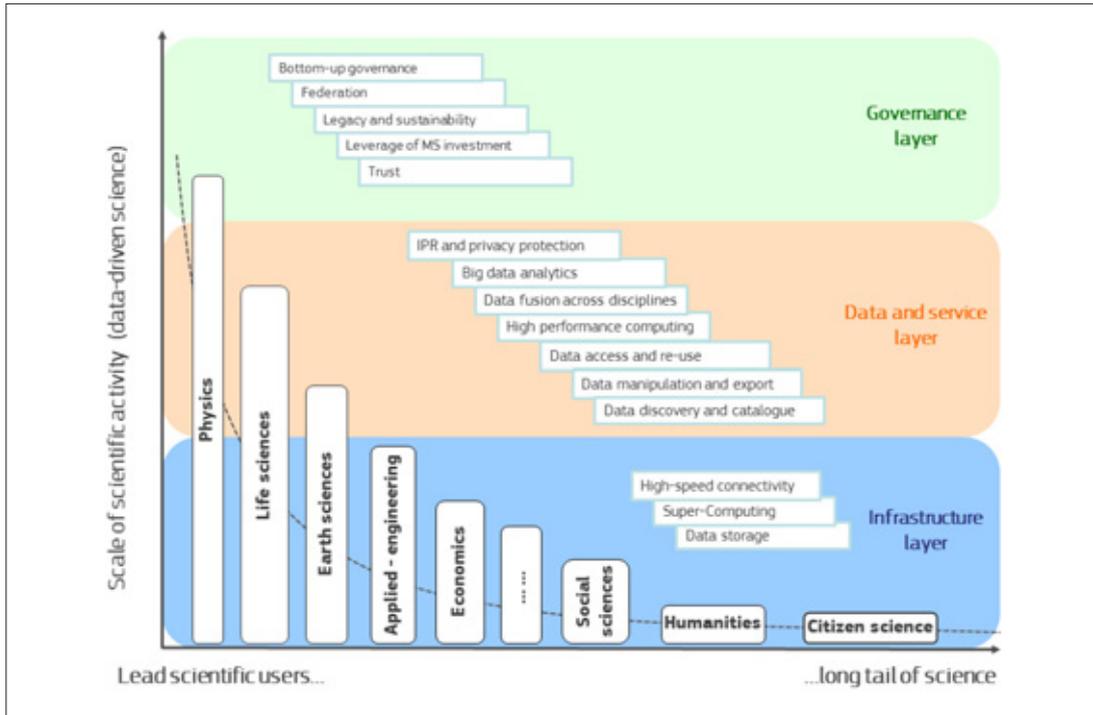
3. 欧州オープンサイエンスクラウド (EOSC) 計画の現状

3-1 欧州オープン科学クラウド計画 (EOSC)

欧州では欧州委員会 (EC) によって 2015 年より始まっている「欧州オープンサイエンスクラウド計画」(European Open Science Cloud：以下 EOSC) が注目を集めている。EC では EOSC を、欧州の 170 万人の科学者、7,000 万人の科学技術専門家のために作られ、研究分野や国境を越えたオープンサイエンスとオープンイノベーションを実現するための基盤であり⁷⁾、「欧州データインフラストラクチャ計画」(European Data Infrastructure) に包含されるものとしている。EOSC は「オープンサイエンス」のための環境であり、この「クラウド」はシームレスな環境や科学データ「コモンズ」を表す隠喩としている。

EOSC においては技術的なチャレンジもさることながら、さらに社会的・制度的なチャレンジも重視される。EOSC は広帯域ネットワーク、HPC、データストレージなどに加えて、データ・レジストリや検索、再利用、ソフトウェアツールや処理系、データ及びメタデータ整備・キュレーション、データの整備と利用をつなぐ専門人材、ポリシー整備、ガバナンスなどを含む包括的システムをコンセプトとしている（図表 1）。

図表1 データ駆動型科学の活動基盤を構成する図



注) 計算機基盤、データに関する整備やキュレーション・サービス、処理系、ガバナンス層がレイヤー構造になっている。データ駆動型科学研究を促進するために、下から上へ向かって、基盤からサービス、データのマネジメント、全体の相互運用性やエコサイクル、ガバナンスが重要となることを示している。(横軸の右から左へ配置されている研究分野は、ECでの議論上象徴的に分かりやすい分野を事例として挙げたものと考えられるが、ここではこれについて議論しない。)

出典：参考文献8

3-2 欧州の社会・経済を変えるデジタル単一市場戦略と EOSC

EUでは2015年5月に、デジタル技術に基づく情報利用・サービス、ネットワークや経済の向上を実現するデジタル単一市場戦略 (Digital Single Market Strategy : DSM) を発表し、5億人が活動する市場に対して、情報通信基盤、雇用創出、公的サービスなどを通じて年間50兆円規模の経済効果が期待できるとしている⁹⁾。ECはDSMにおける社会の利益のためには、データ・情報通信の標準化及び相互運用性 (interoperability) の確保が優先事項であるとしている¹⁰⁾。EOSCはこのDSMの中に位置付けられ、ECの試算によればEOSC構築に67億ユーロ、うち20億ユーロはホライズン2020予算、残り47億ユーロは他の公的・民間資金を併用して投資するとしている¹¹⁾。具体的なEOSCの検討は、2015年9月に設置されたHigh Level Expert Group (以下HLEG) を中心に行われている。HLEG-EOSCは欧州外委員2名を含む10名の外部有識者で組織されており¹²⁾、EOSCイニシアティブの在り方、その戦略などについてECへ提言を行う。EOSC計画は、EOSC構築のための施策・プロジェクト公募を行い、審査の結果採択プロジェクトに予算交付が行われるという枠組みである。本稿執筆時現在、ECからは“European Open Science Cloud

for Research”と題するパイロットプロジェクトの公募が行われている。この公募では2016年度予算で500万~1,000万ユーロを想定するとされている。

3-3 欧州における研究データ基盤の整備事業

欧州内ではEUDAT、GÉANT、LIBER、OpenAIRE、EGIをはじめとして幾つかの研究データ基盤構築に関わる施策、イニシアティブが推進されている。EOSCでは、これらを含む分野的・地理的・施策上別箇に整備されたシステムを結合するとともに、欧州全体の研究データ利活用基盤となることが期待されている。

2015年11月に行われた欧州内ステークホルダーワークショップなど関係者の議論をみると、世界的な共通基盤 (“Global Open Science Cloud”) の欧州セグメントとしてのEOSCを望む声も少なくないようである。2016年3月のRDA第7回総会におけるEOSCの議論では、欧州・米国・豪州以外の地域を含むグローバルネットワーク形成の可能性が示唆された。

また、EOSC関係者において基盤 (インフラストラクチャ) というときに、施設や計算機といったハードウェアだけでは研究データ基盤として機能しないことも専門検討会の席上で言及されている。データキュレータなど専門人材や業務関連の研究開発、サービス業務運用まで全て含めた全体が研究データ基盤として重要

である、という認識はおおよそ共有されているようである。

4. オープンサイエンスをめぐる研究データ共有の取組の国際動向

4-1 データ共有に向けた研究データ基盤システムの必要性

本稿で取り扱うような、オープンサイエンスにおける研究データ基盤構築においては、FAIR原則（FAIR principle：Findable, Accessible, Interoperable, Reusable）¹³⁾ といった相互利用可能な原則に基づいた環境整備が不可欠である。EOSCをはじめとしてこうしたFAIR原則にのっとった研究データ基盤では、多様なデータ形式、データ管理、メタデータ（データに関して記述されたデータ）に対応し、多様な検索、再利用を可能にする体制が必要となる。

こうしたデータ共有や相互運用性のための具体的な課題は、目指すべき研究データ基盤を構築していくための設計理念や指針、あるいはシステムの外部境界条件を形成するものとなり、今後の科学技術データの共有・利用を検討する上で極めて重要であると言える。

4-2 障壁なきデータ利用のための国際組織、RDA

RDAは2013年に開始した国際的な研究データに関する連携を協議する組織で、欧州委員会、米NSF（国立科学財団）、米NIST（国立標準技術研究所）、豪政府らが発足させ、G8国の高級実務者会合下にてきたG8及び関係6か国データ基盤部会との協力の下で運営されてきた。

RDAは国境や専門領域、地域的な障壁のないデータ

共有をはかり、社会的・技術的な研究データ共有・利活用基盤の構築を目的とする。原則として研究者や専門家によるボトムアップ提案で方策、技術などを検討するため、コミュニティによるニュートラルな議論の場を提供するとしている。組織モデルは、インターネットプロトコルのデファクト標準形成に重要な役割を果たしたIETF（インターネット・エンジニアリング・タスクフォース）を下敷きにしてのりである。

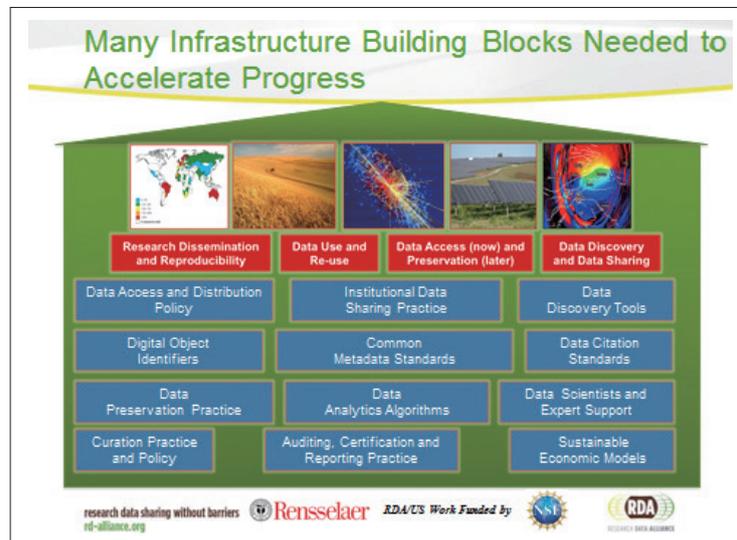
メタデータ、識別子タイプ、データタイプレジストリ、用語定義、データ出版メカニズム、などを課題とした各種の部会（Working Group, Interest Group, Birds of a Feather）がボトムアップに提案され、理事会や技術諮問委員会（Technical Advisory Board：TAB）での審査を経て設置される（図表2）。設置された部会には、興味を持つ個人会員は事実上、誰でも参加可能となっている。このRDAの個人会員はオンラインで簡単に登録でき原稿執筆時においては110か国から4,200名が会員となっている。

RDAの総会は年2回行われており、これまでの開催は欧州域か米国内であった。しかし、2016年3月に開催された第7回総会は初めての欧米外での開催となり、我が国（東京）で開催された。そのため第7回総会では過去数名だった日本人参加者が大幅に増加し、今総会では113名が参加した。

4-3 経済開発協力機構（OECD）とオープンサイエンス

OECD Global Science Forum（GSF）ではオープンサイエンス政策の検討を比較的早くから進めてきた。まず、OECD Principles and Guidelines for Access to Research Data from Public Funding¹⁵⁾

図表2 RDAにおいて検討対象となる、将来のデータ利活用基盤にとって必要と考えられる多くの構成要素



出典：参考文献 14

において、OECD 加盟国及びその他の国が“公的研究で生成された研究データの共有とオープン化を奨励すること”、などを国際社会に対して提言している。これは 2004 年の OECD 加盟国科学技術大臣会合における合意に基づいて検討され、報告書としてまとめられたものである。その後も、社会経済の成長のためには、新たな科学技術の研究開発と、これを通じたイノベーションが今後ますます求められること、これを実現するための方策としてデジタルデータ資源の利活用に基づくデータ駆動型のイノベーションが必要であること、を報告している¹⁶⁾。

その後、2016 年には OECD/GSF がアカデミー系組織 (ICSU-WDS、CODATA) とのジョイントプロジェクトとして、オープンサイエンスのためのデータ基盤の国際連携方策の調査、及び科学データリポジトリの持続的運用可能性についての共同プロジェクト¹⁷⁾を立ち上げている。

4-4 オープンサイエンスに取り組む多くの国際組織と専門家ネットワーク

国際組織の活動においては、国際的な視点で活動できる専門家が重要であり、様々なコミュニティや、イニシアティブの下で設置されたコンソーシアム/国際事業などに関与して調査検討を行ってきた有識者の活躍によるところが大きい。

学術界の最大国際組織である国際科学会議 (International Council for Science : ICSU) が設置した委員会 ICSU-WDS や CODATA をはじめ、他の学術情報やデータアーカイブに関わるコンソーシアム DataCite、DSA (Data Seal of Approval) など、また政府間取組として GEO/GEOSS、環境関連の予算助成機関の連合 Belmont Forum 等で、データリポジトリ・図書館員、データマネジメント、情報学・ICT、各専門領域研究者などが活動している。これらの有識者が上述の RDA や OECD のグループにも加わり実質的な活動主体として議論を行っている。

5. まとめ

5-1 研究データを国際的に相互利用・活用していくために

これまで述べてきたように、欧米をはじめとした国際コミュニティにおいては、オープンサイエンス推進に向けてデータの格納や管理の標準的方法、メタデータの標準化や組織化をはじめとして相互運用性、つまりデータ・インターオペラビリティが重視されている。

相互運用性のある仕組みの構築に当たっては、技術的な側面 (仕様検討やユースケース検討・ツールやレ

ジストリの開発、実装など) と、社会・制度的な側面 (ユーザーアクセスやライセンス、実務シーンにおけるユースケースや問題解決など) とをセットで進める必要があり、そのためにこれまでになかった複合的な専門家集団 (コミュニティ) が必要となる。RDA をはじめとして、本稿に登場する国際組織はこのコミュニティを形成するための核となっていくであろう。

これら、データの技術的・制度的利用の検討、調査というチャレンジに加えて、実際にデータを格納する計算機システムを整備し、ユーザがその上で自由にデータの処理や相互利用を行えるプラットフォームは更に新しい課題として注目される。前述の RDA の設置に寄与した政府機関が、欧州 (EU) では EC の ICT 部局 (EC DG-CNCT) であり、米国では NSF の Cyberinfrastructure 分野であることを考えると、RDA や相互運用性の問題提起の先には、そうした新たな計算基盤や情報通信基盤の将来ビジョンが最初からシナリオとして想定されていた可能性もある。注意いただきたいのは、構築される基盤が十分に活用され成果を生み出すためには、計算機資源や通信基盤、ミドルウェアといった通常のプラットフォームとあわせて、データ管理、キュレーションなどを行う人材の業務開発や育成・雇用、各データの分野ごとの専門性に合わせたソフトウェアサービス、などが不可欠な要素であるということである。

5-2 国際政策としての研究データ利活用推進と日本のプレゼンス向上のために

2013 年 G8 会合での研究データのオープン化合意から 3 年を経て、2016 年の G7 科学技術大臣会合 (茨城県つくば市) では議題の一つにオープンサイエンスが提案された。その中では特に科学研究データの共有、アクセスが中心的な議論となった。採択された「つくばコミュニケ」では、オープンサイエンスに関する作業部会を設置することが明記され、G7 国は OECD、RDA といった科学データ基盤やデータ・インターオペラビリティの国際的議論を行っている組織との連携が望まれることも記載されている。

ここで重要なことは本稿で述べてきたように、将来的に指向されるべき研究データ基盤は、ハードウェア整備や従来型の学術・科学技術情報サービスを越えて、より“総合的な基盤システム”⁴⁾だということである。それは、保存されるデータのキュレーション、サービス、分野横断的な検索、データの再利用を可能とするシステムや必要な標準化、また、分野ごとの高度なデータ整備を可能にする人材の配置やその育成、科学技術や学術研究成果としてのデータに関する評価制度などを含む、新しい意味での研究データ基盤とな

る。上記の G7 科技大臣会合下の作業部会では研究データ基盤の「ハード・技術標準・システム面」「ソフト・人材・制度面」の検討、合意を形成していく議論が必要と考えられる。これは G7 国にとどまらず、国際社会における各国ポリシーに反映され、データ基盤の国際的な相互運用性 (repository interoperability, interoperability of data infrastructures) が推進されるためのステップとなり、さらに将来期待される社会・経済発展のための基盤につながっていくことが期

待される。欧米において基本的な議論や概念形成が進んでいるために日本がその貢献を示すことは容易でないが、一方で、今回の G7 作業部会の議長国が日本と EU の共同となったことは意義深い。今からでも、国際情勢を正確に把握しながら、国内の多くの関係機関・関係者の議論を踏まえて推進することで、G7 国のメンバーとして国際的に先導的な活動に加わることは可能と考える。今後、国内でのより能動的な活動が求められる。

参考文献

- 1) 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」報告書「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：
<http://www8.cao.go.jp/cstp/sonota/openscience/> (access 2016-07-17)
- 2) 村山泰啓、林和弘 (2014)、オープンサイエンスをめぐる新しい潮流 (その1) 科学技術・学術情報共有の枠組みの国際動向と研究のオープンデータ、科学技術動向、146、p.12-17：<http://hdl.handle.net/11035/2972>
- 3) 村山泰啓、林和弘 (2014)、オープンサイエンスをめぐる新しい潮流 (その2) オープンデータのためのデータ保存・管理体制、科学技術動向、147、p.16-22：<http://hdl.handle.net/11035/2990>
- 4) 林和弘 (2015)、オープンサイエンスが目指すもの：出版・共有プラットフォームから研究プラットフォームへ、情報管理、Vol. 58、p.737-744：<http://doi.org/10.1241/johokanri.58.737>
- 5) 例えば、Begley, C. Glenn and Lee M. Ellis (2012), Drug development: Raise standards for preclinical cancer research, Nature, 483, p.531-533：<http://doi.org/10.1038/483531a>
- 6) Artemis Lavasa (2014), Full Summary of the Long Tail Research Data IG Meeting：
<https://rd-alliance.org/groups/long-tail-research-data-ig/wiki/full-summary-long-tail-research-data-ig-meeting.html> (access 2016-06-28)
- 7) European Commission (2016), “European Cloud Initiative to give Europe a global lead in the data-driven economy”：http://europa.eu/rapid/press-release_IP-16-1408_en.htm (access 2016-07-05)
- 8) Jean-Claude Burgelman (2016), Data Sharing Symposium, <https://jipsti.jst.go.jp/rda/> (access 2016-07-17)
- 9) European Commission (2016), “Digital Single Market”：
<https://ec.europa.eu/digital-single-market/en/digital-single-market> (access 2016-07-05)
- 10) European Commission (2016), “Digital Single Market: Economy & Society”：
<https://ec.europa.eu/digital-single-market/economy-society-digital-single-market> (access 2016-07-05)
- 11) European Commission (2016), “Fact Sheet: Digital Single Market – Digitising European Industry Questions & Answers”：http://europa.eu/rapid/press-release_MEMO-16-1409_en.htm (access 2016-07-05)
- 12) European Commission (2015), “Commission High Level Expert Group European Open Science Cloud (E03353)”：
<http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetail&groupID=3353> (access 2016-07-05)
- 13) FORCE11, “Guiding Principles For Findable, Accessible, Interoperable And Re-Usable Data Publishing Version B1.0”：
<https://www.force11.org/fairprinciples> (access 2016-07-17)
- 14) Kathy Fontaine (2015), “RDA Governance”：
<http://www.slideshare.net/KathyFontaine/rda-governance> (access 2016-07-17)
- 15) OECD (2007), “OECD Principles and Guidelines for Access to Research Data from Public Funding”：
<http://www.oecd.org/sti/sci-tech/38500813.pdf> (access 2016-07-02)
- 16) 例えば OECD (2015), “Data-Driven Innovation”：<http://doi.org/10.1787/9789264229358-en> (access 2016-07-02)
- 17) OECD, “Open Data for Science - OECD Project”：
<https://www.innovationpolicyplatform.org/open-data-science-oecd-project> (access 2016-07-02)