

## 第3回 SPARC Japan セミナー2017

「オープンサイエンスを超えて」

# データ駆動型の科学研究エコシステムとしてのオープンサイエンス ー過去のコミュニティ実践事例と日本の視点

村山 泰啓

(情報通信研究機構/ICSU-World Data System)

### 講演要旨



今日、オープンサイエンスは、特に国際的科学的政策文脈では、しばしば研究データエコサイクルのデジタル化、またデジタル基盤上での研究過程や研究アウトプットの出版・共有・配信、といった議論をされることが多い。国際的に期待されているグローバル研究データ基盤やそこで可能となる研究ワークフローを今後設計していくためには、最新の電子情報基盤のツールやノウハウが必須であるとともに、ユーザー（当初は科学者や科学業務に関わる者が主体）からみた実現可能なユースケースに基づく、より上位層のコンセプトが必要となる。過去の研究コミュニティにおける研究データ共有の実践事例は、この設計指針をあたえる上で重要である。

### 村山 泰啓



国立研究開発法人情報通信研究機構ソーシャルイノベーションユニット戦略的プログラムオフィス研究統括（現職）。ICSU-WDS (World Data System) 国際科学委員会 ex officio 委員、日本学術会議連携会員、国立極地研究所南極観測審議委員会委員および重点観測専門部会長、国立国会図書館科学技術情報整備審議会委員を務める。1999-2006年は北極域アラスカにおける上層大気観測日米共同研究計画の日本側リーダーを務めた。内閣府「国際動向を踏まえたオープンサイエンスの推進に関する検討会」委員、欧州委員会「欧州オープンサイエンスクラウド高級専門委員会」委員、首都大学東京非常勤講師、公益社団法人日本地球惑星科学連合理事、京都大学生存圏研究所客員教授、などを歴任。文部科学大臣表彰科学技術賞受賞（2007年）。京都大学工学博士（1993年）。

私がこれまで関わってきたオープンサイエンスは、European Commission や、オバマ政権時代のホワイトハウスなどのデジタル・サイエンス・ベースでサイエンスのやり方を変えていこうという政策的な働き掛けと、コミュニティ側の新しいデジタル時代のオープンネスのノルムや作法、スカラリーコミュニケーションのやり方を変えていくという議論が中心でした。そこにおいて、私の議論の中心は、オープン・アクセス・ジャーナルよりは、むしろデジタルデータにありました。

### 自己紹介

私の研究分野は地球科学、大気科学、宇宙物理で、日本地球惑星科学連合のボードメンバーや、京都大学の客員教授を務めたことがあります（図 1）。最近では、データインフォメーションの活動に日本からも貢献したいと考えて、アメリカ地球物理学連合（AGU）や国際測地学・地球物理学連合（IUGG）といった国際学会の中で委員などを行っています。日本学術会議連携会員でもあります。また、The ICSU World Data System (ICSU-WDS) の国際科学委員会 ex officio 委員で、これがデータについて取り組むきっかけになりました。他に、2016 年まで欧州委員会「欧州オープンサイエ

ンスクラウド高級専門委員会」委員を務め、内閣府「国際動向を踏まえたオープンサイエンスの推進に関する検討会」委員、G7 オープンサイエンス WG のメンバーも務めています。

## 科学のあり方を考える

科学に携わっていると、日本政府でもどこでも、3年か5年のファンドを出すと玉手箱のようにいい成果が出て、例えば災害対策ができるというような政策決定の仕方をよく耳にするのですが、それは正しい科学のやり方なのでしょうか。科学というのは、データを集めて、論考を合理的に重ねていくというロジックの積み上げです。この煩雑な作業を通して、このあたりが確実だろうと言えるものを地道にこつこつ作り上げた結果として、今のテクノロジーがあるのです。スマートフォンにせよ、航空機にせよ、さまざまな現代のテクノロジーはそういったとても面倒な作業の上に成り立っていて、それでわれわれはベネフィットを得られています。

しかし、それを忘れて、何年間か予算を出せば科学者集団が社会に対していいものを出すだろう、金をやるから成果を持ってこいということにどうもなりがちです。不幸なことに、社会と科学のミスコミュニケーションが起っています。それを解決するためにオープンサイエンスという考え方が役に立つのではないかと思っています。私の話はいつも抽象的なのですが、これはなぜかという、この話を政策決定に関わる人

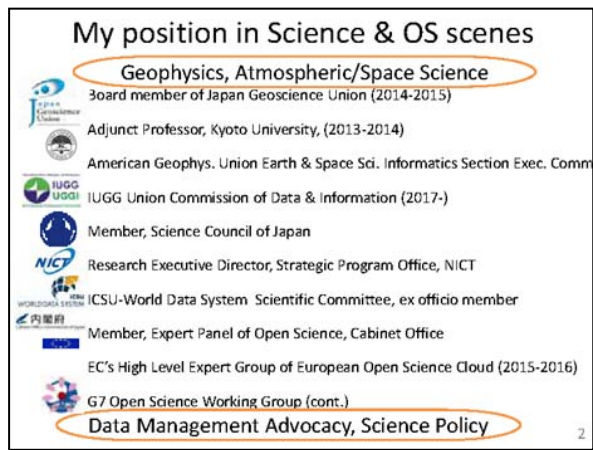
などに、将来の日本の科学のあり方を考えるための基盤にぜひしてほしいからです。

具体的なことでは、データに DOI を付ける、学協会のジャーナルのポリシーをデータに合わせて変えるといった話をするのは重要だと思いますし、そういう話を実際にしている学協会もありますが、そこだけ話しても、何のためにそれをやるのかは恐らく伝わらないと思います。私は SPARC Japan セミナーでも2013年から、もう3~4年この話をしています。日本の中で、科学者は面倒なことをしているし、言うことを聞かないししょうがないけれども、科学に基礎を置いて政治意思決定をする世の中の中でやらざるを得ないのだということを共通理解として持てるようになればと思って、辛抱強く何度も何度もお話をしています。

そういった理由があるので、私はオープンサイエンスについて話すときは、ポリティカルバックグラウンドも必ず話します。その上で、サイエンティフィックリサーチのシステムは、Paul David 氏のお話と根底は同じです。そういうものの国際的なサイエンスノルムの具現化の一つとして、International Council for Science (ICSU) が開いているデータ委員会の一つが World Data System です。科学界だけではなく、社会全体で、デジタルインフラの上で社会のデータまで含めて将来像を変えていこうという動きがあるのです。こういったことを行っていく上で、もちろん具体的に学協会がどう変わらなくてはならないかということもありますが、実際に動きだしているということを少しずつでも議論できたらと思っています。

## オープンサイエンスの国際的な動向

私がオープンサイエンスに関わりはじめたきっかけは、海外の学協会や国際会議で、研究グループ、コミュニティや社会とデータを共有していこうというオープンサイエンスの議論を聞いたとき、何だこれはと思ったことです。当時、全く聞いたことのない話だったのです。その後、G8 科学大臣会合の研究データのオ



(図 1)

オープン化の話が内閣府から来て（図 2）、「こんな議論があったのだけど、あなたがやっている仕事はこれに近いのではないかと」言われて話してみると、まさに同じコンセプトでした。ですから、国際社会では、アカデミーの側とポリティクスの側で、同時にオープンサイエンスが表に出てきた感があります。

当時は、日本でオープンサイエンスの話をしたら、「何でこんな話をするのだ」と言われましたし、いろいろな研究機関の方にお話をしても、「何それ」と引かれるのが関の山だったのですが、あれよあれよという間に重要なトピックになってきたのではないかと思います。

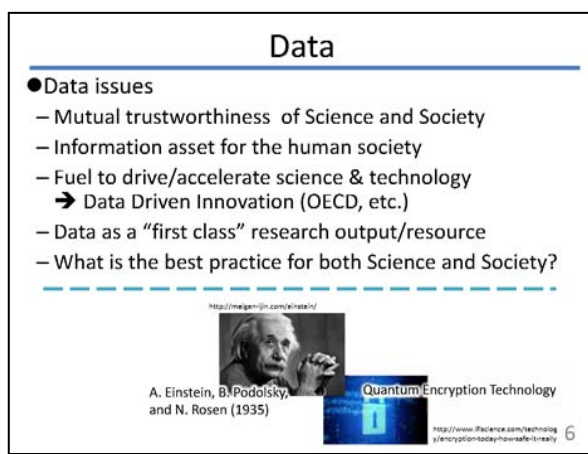
2016年のG7 Science Ministerial Meetingでは、林和弘さんと一緒に話題提供講演をさせていただきました。オープンサイエンスのワーキンググループでは、G7の合意の下に、インセンティブやデータインフラストラクチャーの重要性についての議論が進んでいます。昨年イタリアでG7があり、その次に向けてどうしていくかが考えられています。

でも、何か偉い人たちが集まる場でこういうことを言っているから、オープンサイエンスはきっといいことがあるに違いないと思っても仕方がないのです。研究者や図書館員、出版社、学生、研究機関、企業など、いろいろな人にとってオープンサイエンスは重要だけれど、トップダウンで進めなくてはいけないことも少しはあって、そこがないとどうしてもこれは進まないという部分をきちんとトップが集まる場で決め

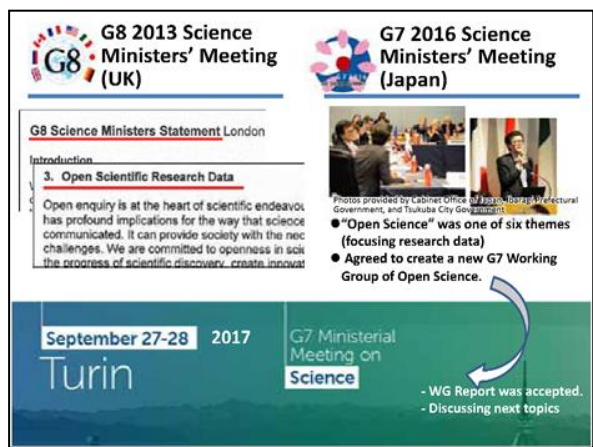
なくてはなりません。そうでないことをむやみにトップが集まる場に持ち込んでも意味がなく、お祭り騒ぎになるだけであって、本当に進みはしないので、そこところは重要だと思っています。国内で議論していると、そこをあまり理解いただけていない場合があるのですが、何でも持ち込めばいいという話ではないのです。

図3は、2016年のG7 Science Ministerial Meetingで出したスライドの一つです。これは単に政治家にメッセージを伝えるためだけのものです。データは社会と科学の相互信頼の問題であって、データはテクノロジーや科学を進めるための fuel（燃料）で、データは研究の成果としても非常に重要な生産物なのだというのを改めてお伝えしました。

日本では内閣府が、「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」を開催し、オープンサ



(図 3)



(図 2)



(図 4)

イエンスのナショナルプリンシプルのようなものをレポート「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」にまとめました（図 4）。ウェブに英語版も日本語版も出ています。これは原山優子議員が Council for Science, Technology and Innovation (CSTI) の代表として作り上げたものです。座長は九州大学の元総長の有川節夫先生で、非常に重要な議論をされたと思います。私もこの委員会に入らせてもらって、作業に関わりました。

### 知識の共有と再現性

本日は出版社やライブラリアンの方が多いと聞いたので、そこの論点も少し入れていこうと思っています。図 5 はその内閣府のレポートの中に出ている図です。当然ながら、内閣府があって、その関連省庁があって、ファンディングエージェンシーがあって、研究機関があるのですが、図書館やリポジトリもきちんと位置付けられています。これは本当に重要なのです。同時に、サイエンスカウンシル、学協会、ジャーナルパブリッシャーが、サイエンスのシステムとして入っています。

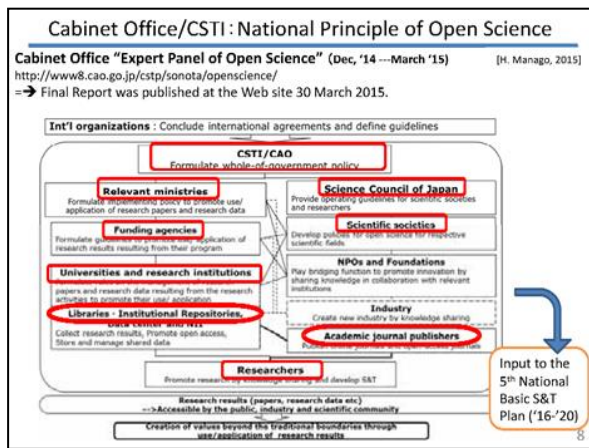
国立大学が比較的メジャーな日本で、ここで省庁から国立大学に話が降りてきて、法的な強制力でもってデータをオープンにするということになると、サイエンティストの振る舞いでは全くなくなります。この、カルチャーを変えていくという部分を抜きにして、強制力が働くのは望ましくない、だからこそ、ここには法的な強制力でないサイエンスソサエティが入ってこ

なければいけないというのが私の意見です。

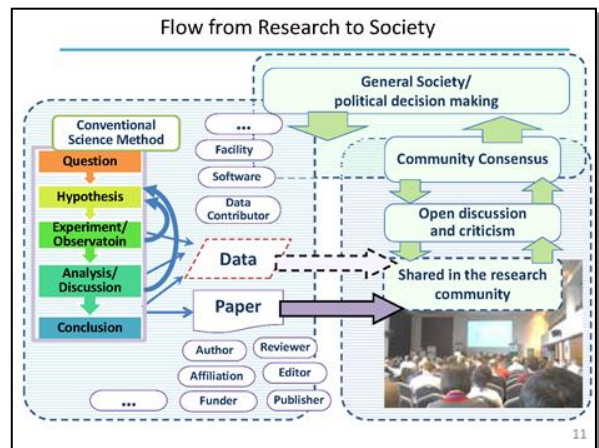
伝統的なサイエンスのメソドロジーは、クエスチョンからハイポセシスが出て、分析してコンクルージョンを得るというもので、これが学生向けのテキストなどに載っています（図 6）。ここから出てきた結論を学協会に発表することが、改めて知識を共有する、サイエンスのオープンネスであり、ディスインタレストッドネス（公平）です。オープンなディスカッションと相互批判をすることで、みんなの共通理解ができていくというプロセスを経ることが重要だということです。そのため、企業の中で閉じて研究するような場合は、みんなの共通理解になりません。それをオープンにして進歩をさらに進めていく上で、論文は非常に重要な役割を果たしてきたのですが、ここで今さらながら、現代の論文というのは検証可能かという話に立ち戻ります。

論文に図を載せて、過去 20 年でこのような気候変動があった、だから次はこの時期が問題だと言うけれど、その図は本当に正しいのか、著者が徹夜して書いたものではないということがきちんとデータでバックアップされなければならないということです。

ここでデータが非常に重要なアクターとなり、データサイエンスという言葉も出てきます。私の感触だと、私が博士学位を取ったのは 1993 年ですが、当時既にデータにものをいわせて博士論文を書いているのです。当時、既にそこではデータサイエンスは成立していたのではないかとというのが私の意見です。それ以前の私



(図 5)



(図 6)

の先生が論文で使っていたデータは、論文中にデータテーブルで書ける程度の量だったのですが、私ぐらいの世代からは、もうそのような量のデータではありませんでした。統計処理して抽象化された図を出すしかなかったのです。そこでモードチェンジが起こっていたのではないかという気がします。

本当はいろいろな付帯物が一緒に出なければ結果は再現できず、データで全て終わりというわけでは全くないのですが、今はデータを最初の第一歩として、データだけにまずは集中してデータの再現性がいわれているという状況です。

### ソーシャルシステムの中の科学

Intergovernmental Panel on Climate Change (IPCC) のレポートは社会に重大なインパクトを与えてきたということは何度もいろいろなところで話しているとおりです (図 7)。1,000 人規模の専門家が集まって一つの報告書を書いています。このようなコミュニティコンセンサスがつくれてこそ、科学的アウトプットは社会にとって重要な知見であり得るし、また、それで実際に政策決定者が影響を受けて社会が変わっていき、社会のベネフィットが端的に表れると言えると思います。

研究活動を行い、それを発表し、蓄積し、引用したり再利用したりして、エコサイクルを回すことで、サイエンスの営みが成立してきたのですが、そこでつくられた科学的理解のコンセンサスが社会に提供されて初めてその有用性が分かります (図 8)。これは、貴

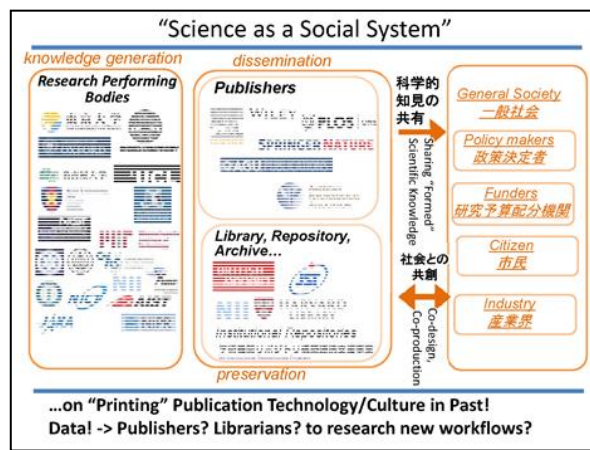
族の時代のように、「この知識はすごいだろう」という話で済むのであれば要りません。いい知識さえ出れば自慢できるというものではなく、一般社会へ伝わらなければ、科学的知見が次の世代に伝わらないし、次の世代で教育されていかないと次の研究が進みません。われわれの科学というのは、研究機関から、出版社・図書館・リポジトリなどを経て、一般社会で使われて初めて、ホビーではなく、社会の営みになっていくのです。

それをする上で、われわれが研究して、出版社・図書館・リポジトリが付いてきてくれればオッケーかというそうではありません。われわれが研究してパブリックファンディングを受けられるように、きちんと一般社会に情報が渡るためには、研究機関と、重要な科学のパートナーである出版社・図書館・リポジトリが共に働ける状況でなければなりません。そういうシステムをわれわれは意識せずにつくってきました。明治に近代科学が日本に輸入されたときに、法律で大学図書館を必ず設置することになったのも、まさに近代科学の方法論を導入した現れではなかったかと思えます。それが今、PDF で論文が取れるからといって、図書館が大学から消えていいのではないかという議論になるのは、サイエンスのオープンネスやフェアネスなど、原理に基づいた思考がアジアの国でなかなかされていないからではないかと思えます。

出版社・図書館・リポジトリは、印刷技術の基盤において育ってきたカルチャーでできているのですが、



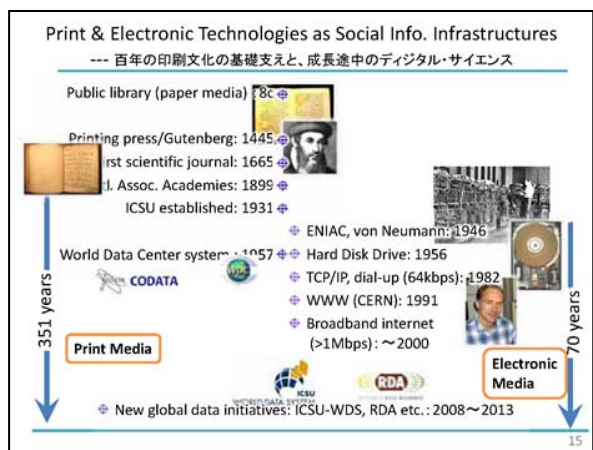
(図 7)



(図 8)

それがデジタルになった瞬間に崩壊したのでは、近代科学の考えそのものが崩壊しますから、それでは大学も研究機関も 10 年、20 年たったら用無しになってしまいかねません。科学を守ることは、われわれ自身を守ることでもあるし、国を守ることでもあるのですが、同時に世界の文化を守ることでもある、そういった意味で、われわれが日本の中で近代科学のやり方をどうやって存続させていくかということが、これから非常に問われる問題ではないかと思えます。

印刷文化は、300 年以上という非常に長い間、科学を進めるための手段として役に立ってきましたが、デジタルメディアというのは非常に短い 100 年未満の歴史の中でここまで普及してきたし、みんなが関心を持って使っています (図 9)。しかし、ソーシャルシステムとして、プリンティングテクノロジーと同じぐらい信頼できるのかどうかというのはこれから問われる問題で、そこをわれわれが今解決しなければいけません。研究者はデジタルメディアを使って、テクノロジーを今までつくってきたのですが、それを社会基盤にしていくところで、研究者だけではもう立ちいかなくなったのです。だからこそ Research Data Alliance のように、プラクティショナー、ライブラリアン、いろんな人が入ってきて、ポリシーメーカーも入ってきて、一緒にソーシャルインフラストラクチャーにしていくためにどうしたらいいか、国際的にみんなが悩んでいます。その中で、アカデミーからも World Data System など国際組織もいろいろ入って議論しているのです。



(図 9)

WDS の歴史は前史からいくと 60 年ぐらいあるのですが、今のデジタルインフラストラクチャーに対応するためにいろいろなことを変えなければいけないし、元々地球科学が中心だったのですが、メディカルからエンジニアリング、ソーシャルサイエンスまでいろいろなものが今入ってきて、データ基盤をどうするか考えているところです。

データを蓄積するために、CoreTrustSeal という新しいデータリポジトリ認証の仕組みを WDS とヨーロッパの Data Seal of Approval という機関で立ち上げました。要するに、サイバーセキュリティのような信頼性と別に、データそのものの信頼性を長期間担保するために、国際枠組みが必要だと考えて、そのための取り組みを行ってきたのです。

図 10 はアメリカ国立衛生研究所 (NIH) のホームページに載っていた図で、ISO と WDS はそれぞれレベルの違うサーティフィケーションなので、国際認証がなくてはいけないという話をしているということです。

そういったことを具体的に社会の中でインプリメントするために、European Commission は digital single market strategy、日本政府は Society 5.0 を考えているのですが、やるとしたら 10 年、20 年、30 年かかり、カルチャーもインダストリーもエデュケーションもみんな変わっていかないといけないのだから、5 年間の five year strategic plan で完成する話ではないと思います (図 11)。それはインターネットの世界でも当然

Figure 10 is a table titled "International Framework of 'Trustworthy Data Repository' Certification". It compares three certification levels: CORE, EXTENDED, and FORMAL. The table includes columns for LEVEL, CORE, EXTENDED, and FORMAL, and rows for various criteria.

LEVEL	CORE	EXTENDED	FORMAL
Organization(s)	WDS: ICSU World Data System DSA: Data Seal of Approval	DIN: German Institute for Standardization	ISO: International Organization for Standardization
No. of Requirements	16	34	100+
Standards	Mandatory Requirements	DIN 31604	ISO 14721 (DAIS) ISO 15363 ISO 15919 ISO 17021
Audit Process	Self-assessment + independent peer review (2)	Self-assessment + independent peer review (2)	ISO certified audit with accredited auditors
Certification Cost	Free	Charged after Jan. 2018 €500	\$10,000
Designation	World Data System Wagon or Data Seal of Approval	Trustworthy Digital Archival	TBD
Certification lifespan	3 years	Indefinite	3 years
No. of Certified Repositories	130+ (WDS)   DSA ( )	2	Coming Soon

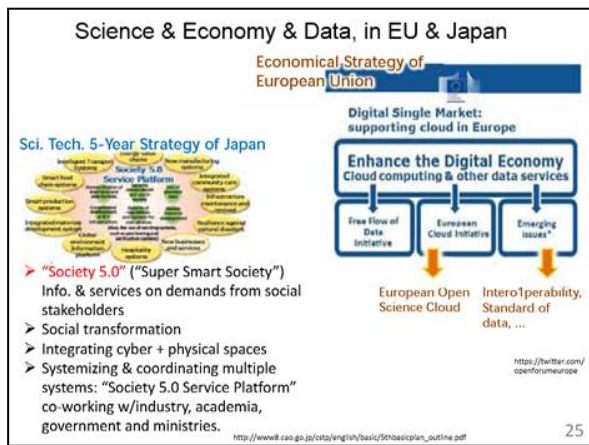
Source: [https://datascience.nih.gov/Trusted\\_Digital\\_Repository](https://datascience.nih.gov/Trusted_Digital_Repository)

(図 10)

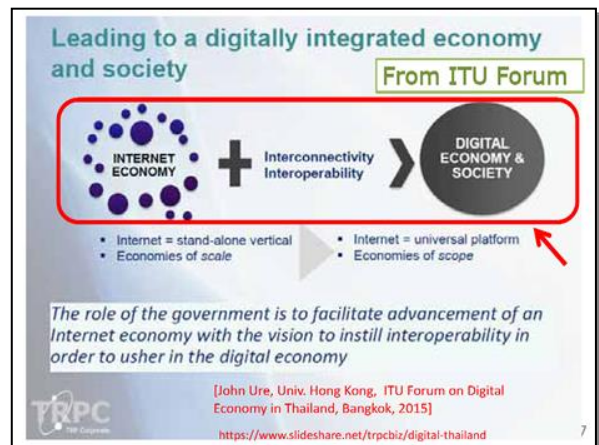
議論されています。

インターネットエコノミー+インターコネクティビティ、インターオペラビリティで、デジタルエコノミーになるというのは、ITU Forum、情報通信の世界でも議論されています（図 12）。

研究者は、ファンディングや省庁を通じて政治的な合意の影響を受けますが、そこに通底して国際的な共通理解がなければ、サイエンスとしては機能しません。それを機能させていくために、リサーチ・アウトプット・マネジメントという形で、図書館・出版社・アーカイブ・リポジトリといったものが、単なる研究のサポーターではなく、社会との関わり全てを支える骨組みとして、研究者と共に生きていってほしいと思います。これらが不可欠なパートナーとして存在し続けていくことが、科学そのもの、現代文明そのものの基盤になるのだろうという枠組みを描きながら、日本の将来が議論できるといいのではないかと考えています。



(図 11)



(図 12)