オープンサイエンス政策と研究データ同盟（RDA）が進める研究データ共有と、デジタルアーカイブの接点に関する一考察：新しい研究パラダイムの構築に向けて

科学技術・学術政策研究所
科学技術予測センター
林　和弘

デジタルアーカイブ学会第2回研究大会
東京大学
2018.3.10

CC BY

1はじめに

デジタルアーカイブとその利活用

- ICT（Information and Communication Technology）の活用
  - 文化財の保存（アーカイブ）に加えてその利用も効率化されることとなった。
  - 人文社会学の新しい研究活動
    - みんなで翻刻　https://honkoku.org/
    - 江戸料理レシピ　http://codh.rois.ac.jp/edo-cooking/
    - 社会への波及効果の新たな展開をも生み出している。
- 世界の科学技術・学術政策（Science and Technology Policy）
  - オープンサイエンス政策と様々な取組
  - その根底にある考え方
    - ICTを活用したデジタルネイティブな新しい研究活動を志向
    - これまでの電子ジャーナル、論文、被引用数で構成されるものとは根本的に異なる
    - 連動する産業を生み出すことも念頭に置いている

2. オープンサイエンス政策と研究データ連盟（RDA）の取組

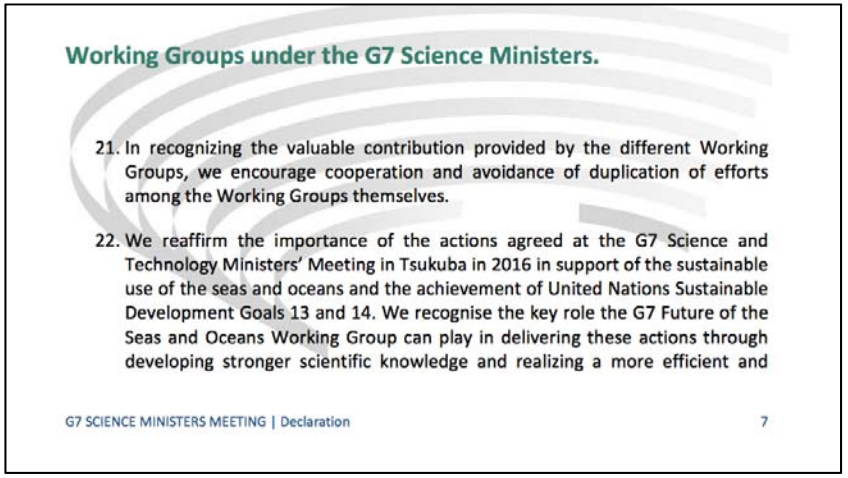
2.1 新しい研究パラダイムを志向するオープンサイエンス政策

- 世界のオープンサイエンス政策
  - 2013年：G8ロンドン・サミット
    - オープンデータの議論が端緒
  - 2016年：つくば（日本）科学技術大臣会合
    - 初めてオープンサイエンスが議題に挙がり、オープンサイエンスの作業部会が設置
  - 2017年：トリノ（イタリア）科学技術大臣会合
    - トリノ会合での報告書（コミュニケ）においては、オープンサイエンスは、新しい研究パラダイムへの変容を志向していることが述べられている。
    - 作業部会は引き続き継続
  - 欧州連合（EC）を中心にトップダウンの政策に基づき、研究データ基盤整備など、その実装が検討されている。
  - 良くある間違い：研究論文やデータをすべてオープンにする活動

G7 Italy (Trin)

- **2017年9月のG7科学技術大臣会合（イタリア）**
  - **オープンサイエンス作業部会の報告書の披露**
- **オープンサイエンスは新しい研究パラダイムを生み出すものとして作業部会が継続することになった。**





address complex economic and societal challenges, are transforming the way in which science is performed **towards Open Science paradigms**.

http://www.g7italy.it/sites/default/files/documents/G7%20Science%20Communiqu%C3%A9.pdf

2.2 オープンサイエンスの実践と試行錯誤の場であるRDA

- **RDA（Research Data Alliance：研究データ同盟）**
  - 2013年：ロンドンG8科学技術大臣会合の共同声明
    - 2013年3月：「障壁なき研究データ共有」をスローガンとして結成、オープンサイエンス政策の実践の先頭に立つ組織
  - 設立メンバー
    - EC（欧州委員会）
    - NSF（米国国立科学財団）
    - NIST（米国国立標準技術研究所）
    - ANDS（オーストラリア国立データサービス）、
    - オーストラリアイノベーション省である。
  - 2016年3月　日本(東京)で第7回総会開催
    - アジアとしては初めて



2.2 オープンサイエンスの実践と試行錯誤の場であるRDA

- RDAの特徴：<u>ボトムアップ＆マルチステークホルダー</u>
  - 研究者を中心とした科学の発展のための具体的な事例に基づくデータ利活用の議論
    - これは元からある、ICSU-WDS、CO-DATA等
  - データ出版、データ引用と研究評価、データの質の保証の仕方や、データの相互運用性を法律的にどう担保していくかなども議論
  - 研究データ基盤の運用や研究者のキャリアパス形成を含めて、研究活動全体（エコシステム）の改善や研究評価を含む研究者社会の発展に関する議論も
  - 研究者に加えて出版者や図書館、研究助成団体など多様なステークホルダーが参画

2.2 オープンサイエンスの実践と試行錯誤の場であるRDA

- 実践をより重視
  - セッション構成
    - BoF(Bird of Feather)：関心を持つ人々の集まり
    - IG(Interest Group)：提言等まとまりのあるイシューの探索
    - WG(Working Group)：課題と提言をまとめる
  - WGになったセッションは一定期間（18ヶ月）内に提言を出すことが義務→ECの標準化文書等に反映も
- 多様なステークホルダーが自発的に集まり、
  - 研究データの共有に関する個々の問題を解決する
  - 新しい研究パラダイムに移行するために必要な課題そのものを見出す
- ボトムアップ組織

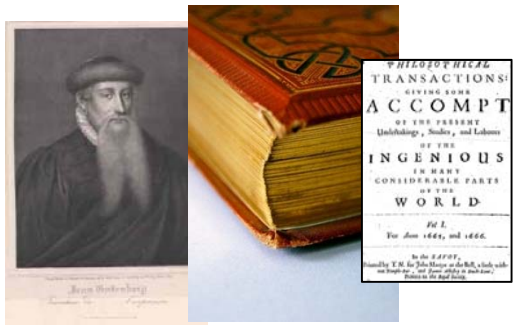
研究文化の再構成

- トップダウンのオープンサイエンス政策とボトムアップによるRDAの活動が当面目指しているのは、研究データとその基盤（プラットフォーム）を念頭においた新しい研究活動の開発
- そのプラットフォームが報酬（研究費）、昇進、名誉と連動し、新しい研究文化が醸成されることを想定
  - 現在の学術ジャーナルが論文という形態とともに17世紀から350年ほどかけて培ってきた研究文化を再構成するもの

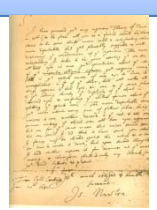
歴史に習う

- ポストグーテンベルグ時代の再フレーム

大量印刷ベース



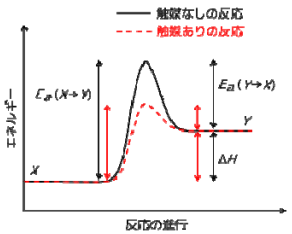
Letter based dissemination



Past Design



遷移状態



http://commons.wikimedia.org/wiki/File:Activation_energy_ja.svg

Web ベース



Future Design



学術情報流通の再発明を含む社会変容

K. Hayashi, “Current States of Impact Assessment of Research Outputs in Japan and Some Challenges to Measure New Impacts for Japan's Stakeholders,” OECD-ESTONIA WORKSHOP ON IMPACT ASSESSMENT: PRACTICES, TECHNIQUES AND POLICY CHALLENGES, May 15-16 2014, Estonia. (revised)

FAIRデータ科学

- 研究データと一口に言っても様々な議論
- FAIR原則に従った研究データを中心とした研究活動基盤の構築
    - Findable: 見つけやすい
    - Accessible: アクセスし易い
    - Interoperable: 相互通用性がある
    - Reusable: 再利用可能である
  - FAIRデータ科学とも言える枠組みを模索
  - （FAIR原則には日本のBioscience研究者も寄与）
- 信頼のあるデータ流通サービスを実現できるか
  - 例えば、データリポジトリの認証なども始まっている。CoreTrustSeal

EOSC Declaration

RECOGNISING the challenges of data driven research in pursuing excellent science;

GRANTING that the vision of European Open Science is that of a research data commons, widely inclusive of all disciplines and Member States, sustainable in the long-term,

CONFIRMING that the implementation of the EOSC is a process, not a project, by its nature iterative and based on constant learning and mutual alignment;

UPHOLDING that the EOSC Summit marked the beginning and not the end of this process, one based on continuous engagement with scientific stakeholders, the European Commission,

PROPOSES that all EOSC stakeholders consider sharing the following intents and will actively support their implementation in the respective capacities:

Data culture and FAIR data

- [**Data culture**] European science must be grounded in a common culture of data stewardship, so that research data is recognised as a significant output of research and is appropriately curated throughout and after the period conducting the research. Only a considerable cultural change will enable long-term reuse for science and for innovation of data created by research activities: no disciplines, institutions or countries must be left behind.

- [**Open access by-default**] All researchers in Europe must enjoy access to an open-by-default, efficient and cross-disciplinary research data environment supported by FAIR data principles. Open access must be the default setting for all results of publicly funded research in Europe, allowing for proportionate limitations only in duly justified cases of personal data protection, confidentiality, IPR concerns, national security or similar (e.g. 'as open as possible and as closed

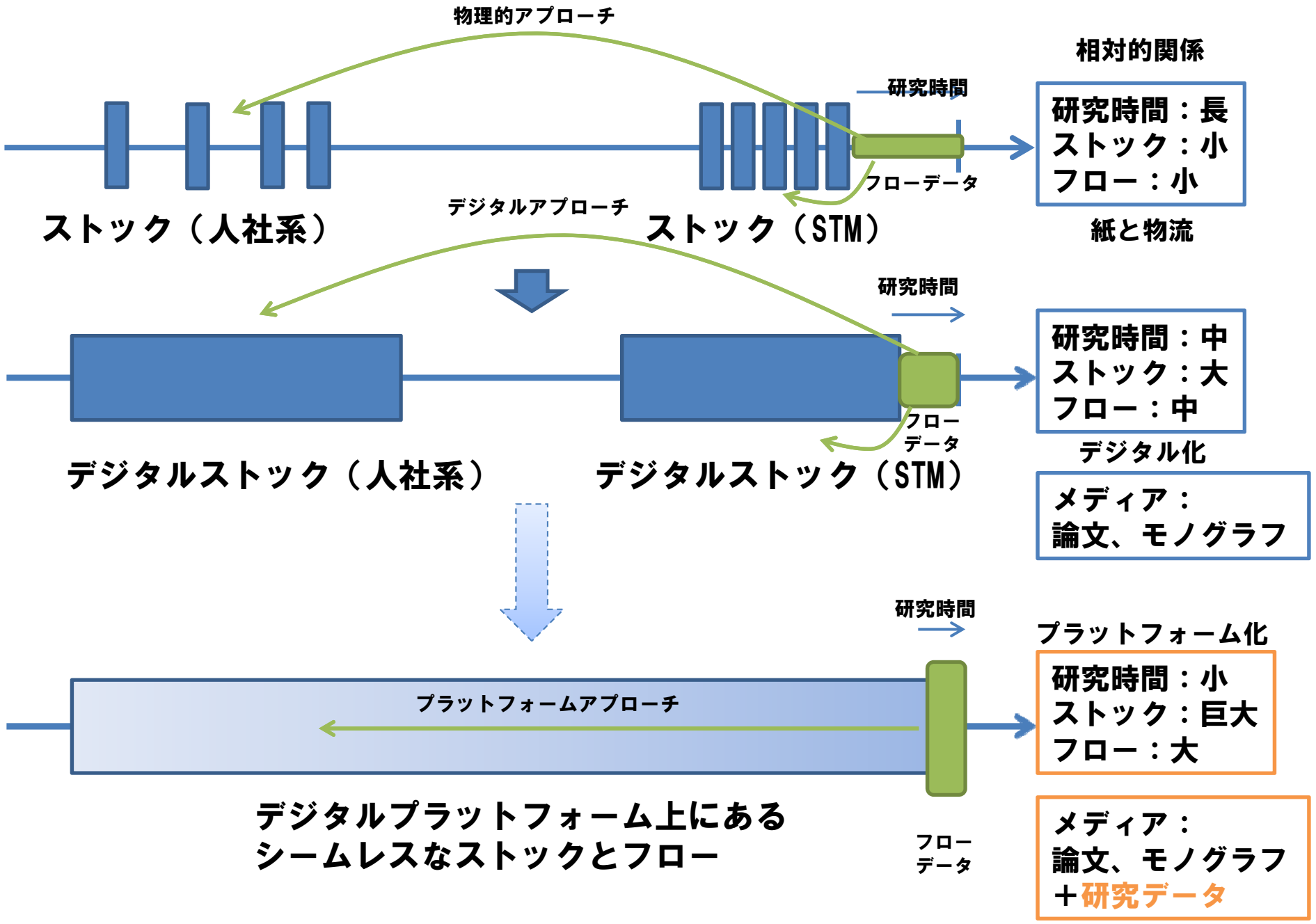
3. オープンサイエンスとデジタルアーカイブの違いと接点

3.1 フローを意識する研究データとストックを意識するデジタルアーカイブ

- RDAとデジタルアーカイブの活動はそれぞれ独立しているように見える
  - デジタル化の対象の違い
    - デジタルアーカイブは、主に文化財のデジタル化に主眼を置いたもの
    - 元々文化財として時に1000年を越えて蓄積（ストック）された情報を扱う
    - RDAが主な対象としているのは、研究者が日々の活動で扱うデータ（フロー）であり、保存（ストック）の議論も数年から数十年レベルであることが多い
  - 研究成果発信媒体の違い
    - 理工医学系が学術ジャーナルの論文が中心
    - 人文社会学系はモノグラフに重きを置く分野も多い。
    - この成果発信スタイルを支えて来たのが紙に情報を印刷する仕組みと、それを郵送する手段（社会制度）

3.2相互通用性のある データフォーマットの重要性

- 超分野横断データフォーマットとしての、論文やモノグラフ
  - 現在（＝紙と物流に依拠した社会制度上）もっとも相互通用性の高い研究データフォーマットと言える。
- 次世代長文や横断データフォーマットwith ICT
  - 新しいデジタルネイティブフォーマットの開発
  - ICTを活用した研究データの流通システムに依拠
  - 時間をかけて研究者コミュニティに受け入れられることは想像に難くない。

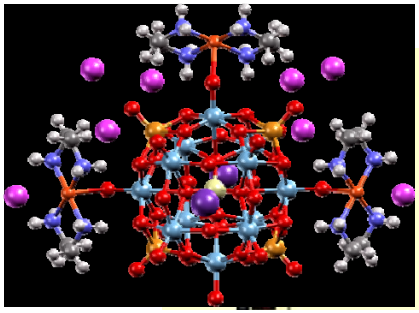

物理的アプローチ
相対的関係
研究時間
研究時間：長
ストック：小
フロー：小
フローデータ
ストック（人社系）
デジタルアプローチ
ストック（STM）
紙と物流
研究時間
研究時間：中
ストック：大
フロー：中
フロー
データ
デジタルストック（人社系）
デジタルストック（STM）
デジタル化
メディア：
論文、モノグラフ
研究時間
プラットフォーム化
研究時間：小
ストック：巨大
フロー：大
プラットフォームアプローチ
デジタルプラットフォーム上にある
シームレスなストックとフロー
フロー
データ
メディア：
論文、モノグラフ
＋研究データ

3.2相互通用性のある データフォーマットの重要性

- データフォーマットがすでに研究データの流通と共有の文化作りに役立っている例
- CIF (Crystallographic Information File)
  - 結晶学のX線構造解析データフォーマット
  - 国際結晶学連盟(International Union of Crystallography: IUCr)にて1990年代に開発された
  - 結晶構造データに加えて実験・計算条件なども含めることができる。
  - 新たな記載事項が生じた際も簡単に追加出来る柔軟性・拡張性が高いデータフォーマット
  - 結晶学に限らず周辺の領域にある分野の論文においても標準として使われ、ケンブリッジ結晶学センターには、多くのCIFファイルが蓄積

結晶学におけるCIFファイル

CIFというファイルフォーマットを中核とする、研究データ共有と論文出版のフローが確立している。



https://www.ccdc.cam.ac.uk/Community/LOFHOF_6.png

http://slideplayer.com/slide/798419/

- **CIF= an interoperable standard format for Chemists**

3. 2相互通用性のある データフォーマットの重要性



- **デジタルアーカイブにおける**IIIF (International Image Interoperability Framework)**の重要性**
  - **画像へのアクセスを標準化し相互運用性を確保するための国際的なコミュニティ活動**
- **画像情報だけでなく、研究活動に役立つ情報も取り込まれようとしている→** CIF**ファイルと同じ流れ**
- **ファイルフォーマット駆動型研究の進展**
  - **研究活動の（現実的な）再構成の可能性**

4. オープンサイエンス時代の新しい研究パラダイムに向けて

新しい研究パラダイムに向けて

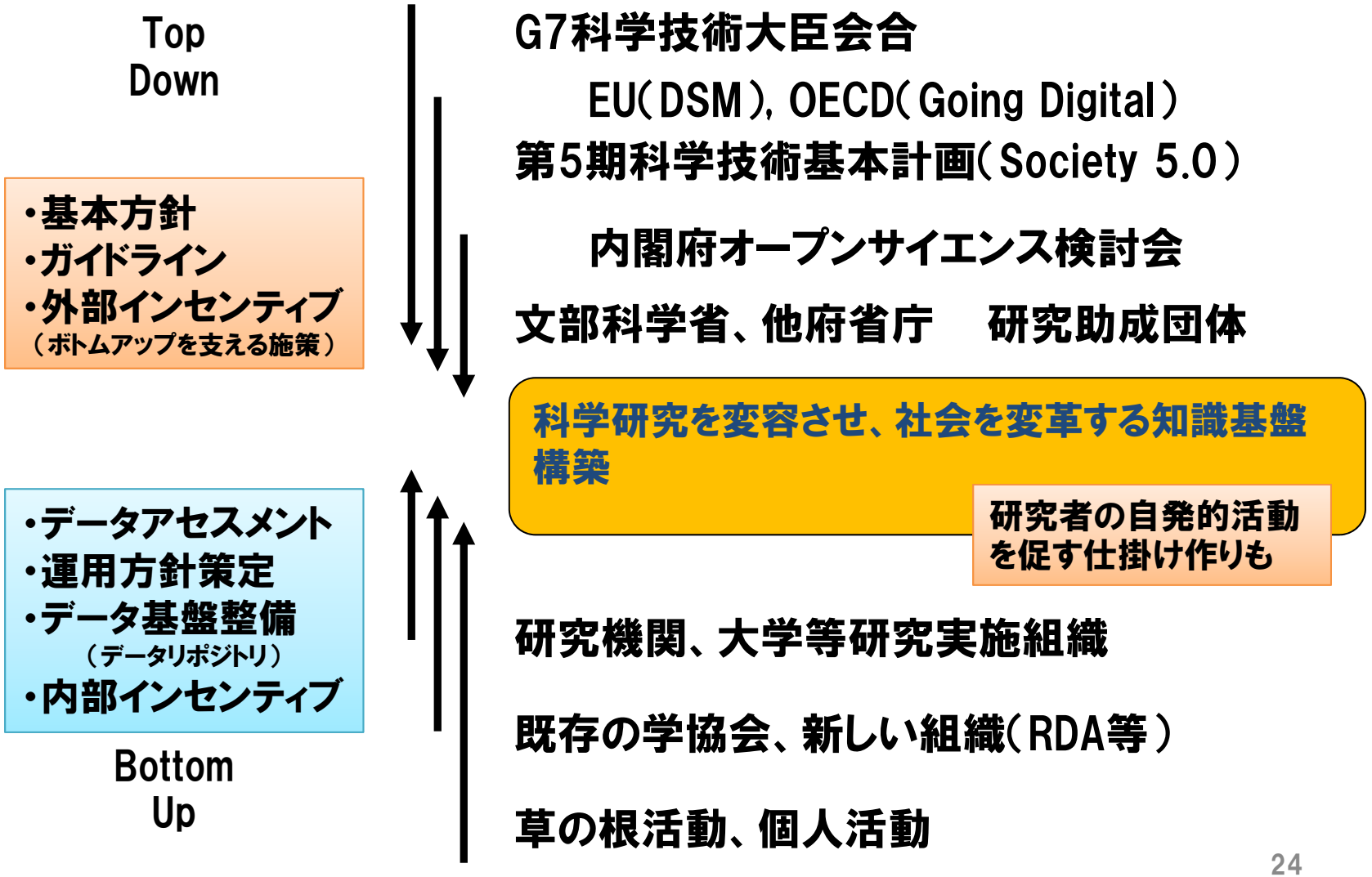
- オープンサイエンス政策が志向する新しい研究パラダイム
  - 文理融合の加速
  - デジタルネイティブの研究基盤（ストック&フロー）
  - デジタルネイティブな研究文化作りの進展
- 必要とされる
  - 将来ビジョンと戦略
  - 信頼された手法や信頼された組織に基づく、研究データ流通サービスの試行錯誤
  - マルチステークホルダー（研究者、出版者、図書館、研究助成団体など）による既存の枠組みを超えた対話と試行の繰り返し
  - RDAで進行し、デジタルアーカイブにおいても必須

信頼をどのように得るか

- **過渡期において研究者を中心としたアクターの信頼をどのように確保するか**

	漸次的な対応 （既存の信頼を得ている手法の拡張）	非連続的な対応 （実行する組織や取組の信頼性に依拠して活動）	解説
論文出版	ドメイン別プレプリントサーバー＋ピアレビュージャーナル（arXive＋物理系ジャーナル）	非政府系研究助成団体のオープンリサーチプラットフォーム（Wellcome Opren Research）	プレプリントとして、論文の草稿をオープンに掲載するのは同じだが、その後の質の保証を既存のジャーナル出版に依拠するか新しい査読の手法を取り入れるか。
データ出版	データジャーナル＋引用（Scientific Data＋Data Citation）	データリポジトリ＋altmetrics（figshare+altmetric.com）	いわゆるジャーナル出版の仕組みとインパクト評価をデータ出版に適用するか、データリポジトリベースにするかの違い。
研究データ共有基盤	機関リポジトリへの研究データ登載＋横断検索(IR+JAIRO Cloud)	研究データ登載・管理プラットフォーム(COS Open Science framework)	論文のリポジトリの拡張によるデータ登載か、データを中心とした管理プラットフォームかの違い。

オープンサイエンスの推進：当面必要とされるトップダウンとボトムアップによる知識基盤構築



まとめ

世界の潮流

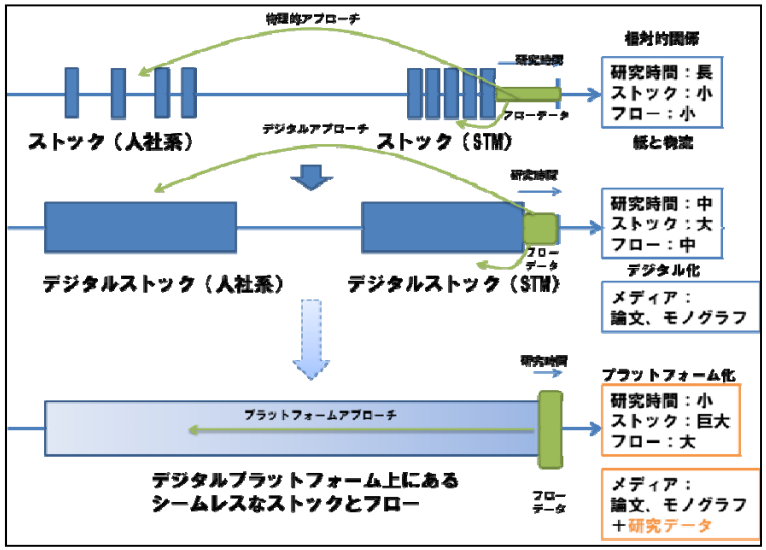


トップダウンとボトムアップ

制度の変革



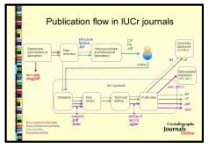
対話と試行錯誤



ストックとフローのシームレス化とプラットフォーム

	漸次的な対応 （既存の信頼を得ている手法の拡張）	非連続的な対応 （実行する組織や取組の信頼性に依拠して活動）	解説
論文出版	ドメイン別プレプリントサーバー＋ピアレビュージャーナル（arXive＋物理系ジャーナル）	非政府系研究助成団体のオープンリサーチプラットフォーム（Wellcome Opren Research）	プレプリントとして、論文の草稿をオープンに掲載するのは同じだが、その後の質の保証を既存のジャーナル出版に依拠するか新しい査読の手法を取り入れるか。
データ出版	データジャーナル＋引用（Scientific Data＋Data Citation）	データリポジトリ＋altmetrics（figshare+altmetric.com）	いわゆるジャーナル出版の仕組みとインパクト評価をデータ出版に適用するか、データリポジトリベースにするかの違い。
研究データ共有基盤	機関リポジトリへの研究データ登載＋横断検索(IR+JAIRO Cloud)	研究データ登載・管理プラットフォーム(COS Open Science framework)	論文のリポジトリの拡張によるデータ登載か、データを中心とした管理プラットフォームかの違い。

研究データ文化醸成と信頼づくり



CIF　IIIF

データフォーマットの重要性

デジタルネイティブ＝文理融合→境界のリフレーム
新しい研究パラダイム構築
オープンサイエンス

