

研究データ利活用に関する国内 活動及び国際動向について

武田英明

国立情報学研究所 教授
研究データ利活用協議会 会長

takeda@nii.ac.jp

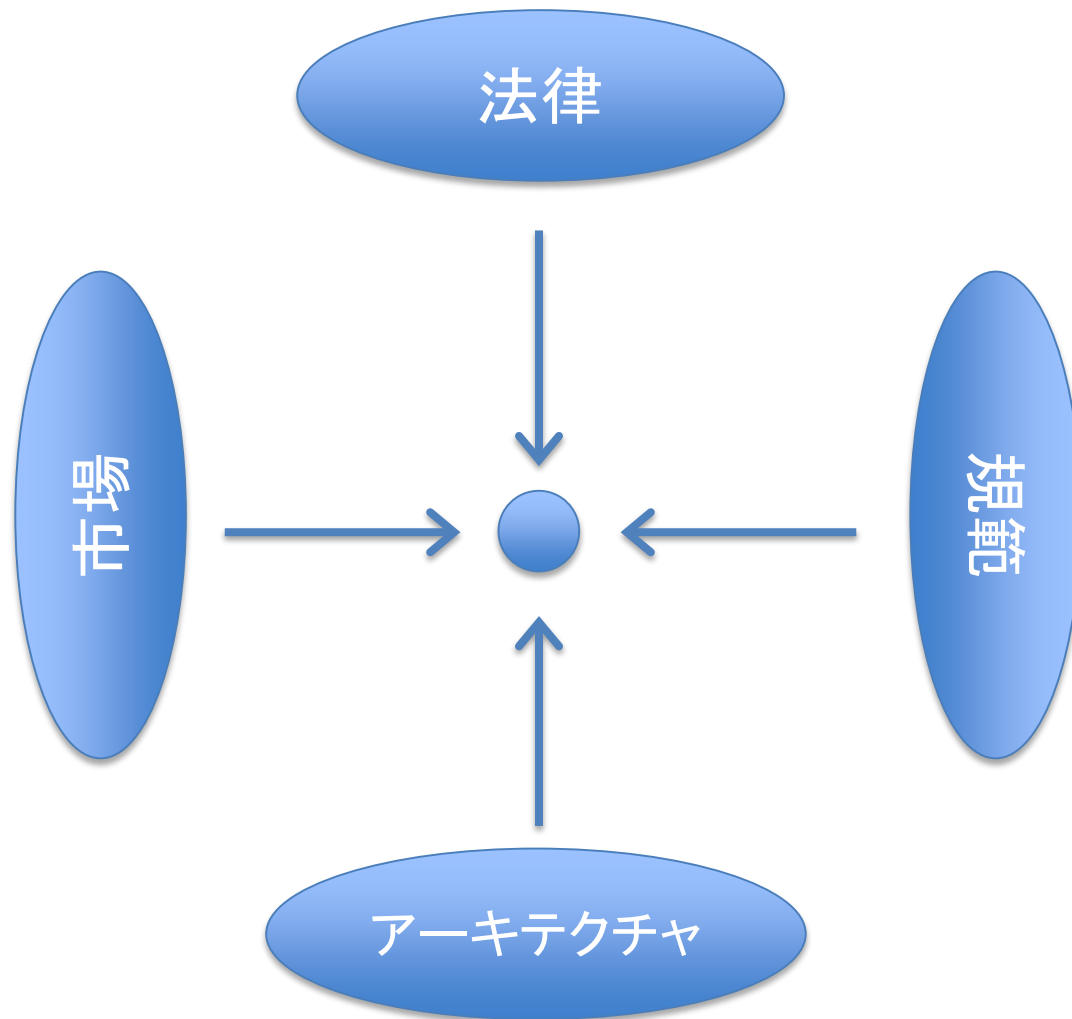
@takechan2000

目次

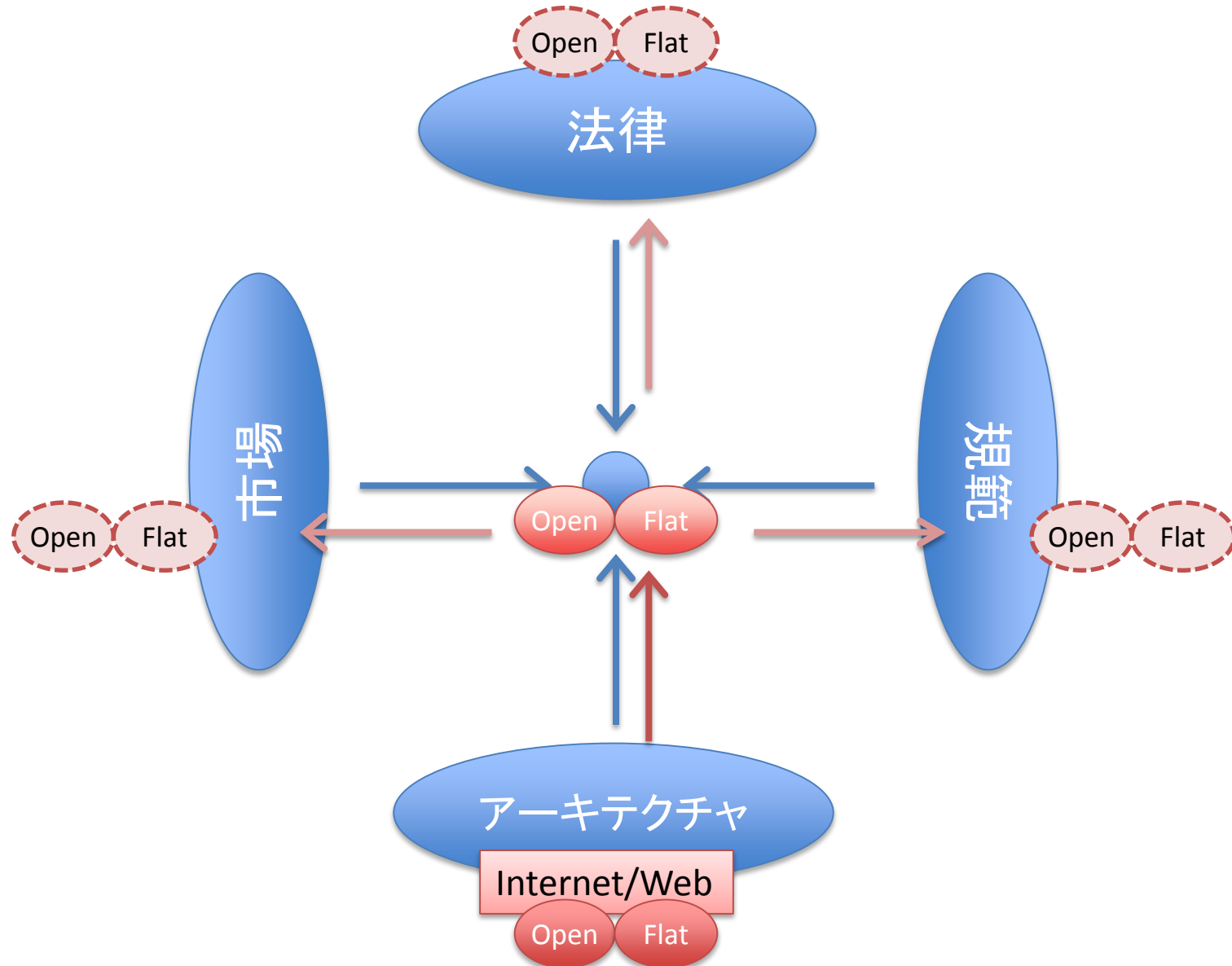
- なぜ今、オープンサイエンスなのか
- オープンサイエンスの系譜
- オープンサイエンスと研究データ共有
- 研究データ共有の枠組み
- 研究データに関わる活動
 - 国際: Research Data Alliance (RDA)
 - 国内: 研究データ利活用協議会 (RDUF)

なぜいまオープンサイエンスなのか

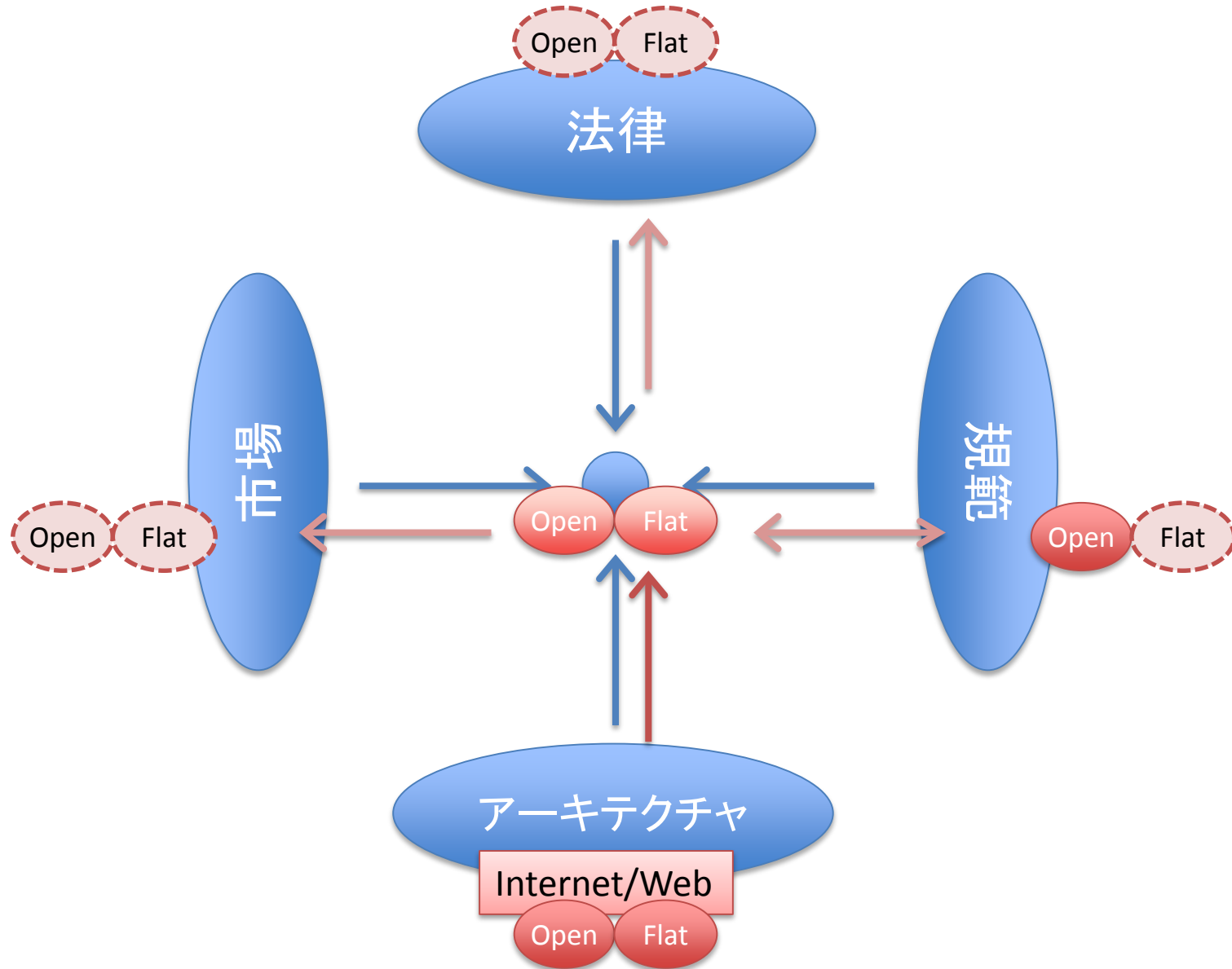
社会における個人に対する4つの規制の様相



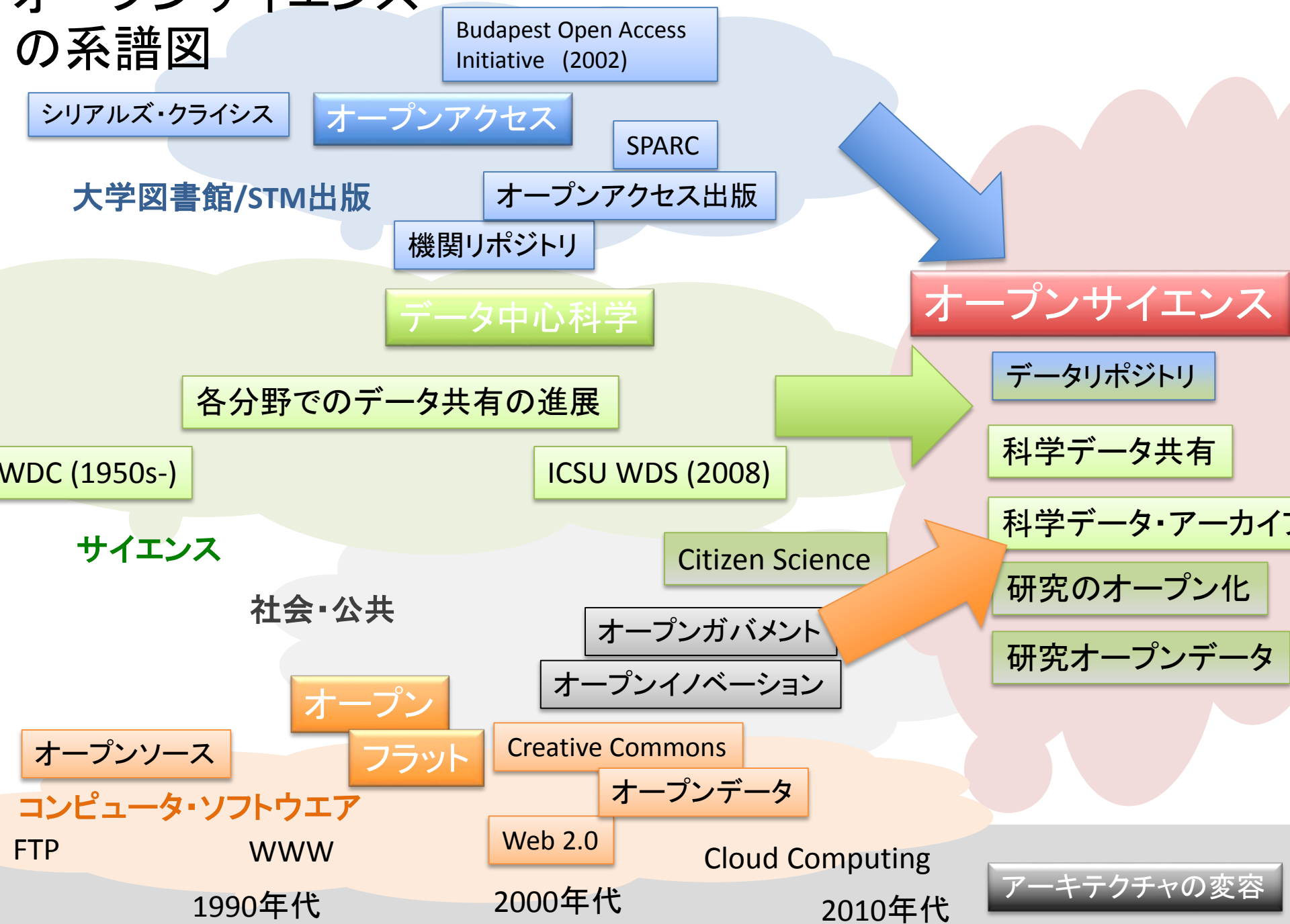
Internet/Webによるアーキテクチャの変容



サイエンスの変容



オープンサイエンスの系譜図



オープンアクセス

- オープンアクセス以前
 - コミュニティ独自の論文共有の世界
 - Preprint共有の伝統(物理学など)
 - -> LANL preprint archive (1991) -> arXiv.org (1999-)
 - Technical reportの刊行
 - 紀要の刊行
 - 論文請求の手紙

オープンアクセス

- シリアルズ・クライシス

- 大学図書館における購読雑誌の減少

- 雑誌購読料が高騰

- 1986年から1999年の間で、雑誌支払単価は207%増加、支払額170%増加、タイトル数6%減少(北米研究図書館協会の調査)[1]

- 対策

- SPARC (STM出版社の非競争環境を変える)
 - オープンアクセス

[1] Case, M. M. (2002). “Capitalizing on Competition: The Economic Underpinnings of SPARC” . SPARC.

オープンアクセス

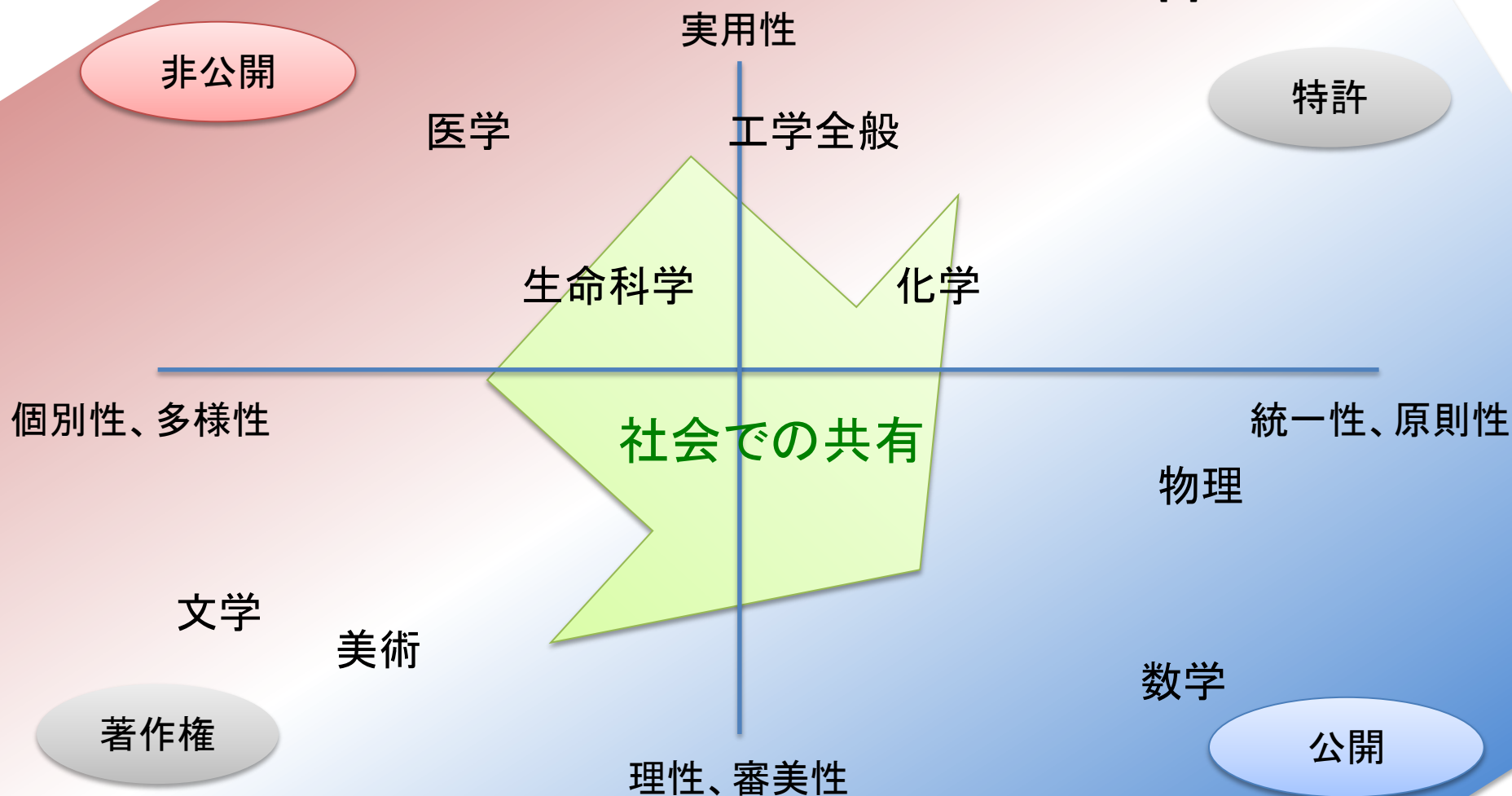
- 論文などの学術情報を無償で自由に利用できるようにすること
 - Budapest Open Access Initiative (2002)
- 理由
 - 論文の可視化とインパクトの最大化(他分野でも)
 - 広範な人からのアクセス(途上国の研究者でもアクセスできる)
- 実現方法
 - セルフアーカイブ(グリーン・ロード)
 - プリプリントサーバ
 - 機関リポジトリ
 - オープンアクセスジャーナル(ゴールド・ロード)
 - APC
- 学術情報
 - 査読付き論文
 - そのほか論文
 - ソフトウェア、データ
- 利用の程度
 - アクセスして閲覧できる
 - 利用や加工ができる

科学でのデータ共有

- 分野ごとのデータ共有
 - 天文学
 - 素粒子物理学
 - 生命科学
 - 地球惑星科学
 - 生物多様性
 - 社会科学
 - ...
- 分野ごとにデータ共有の特性が違う
 - 共有／公開
 - データ量
 - 集中／分散

分野による情報・データの共有可能性

[1]を元に筆者が改変



[1] 有田正規: バイオインフォマティクスの現状とデータシェアリングの可能性について、「データシェアリングを利用した科学技術」に関する勉強会 開催記録 第2回 2015/4/1, 文部科学省・科学技術振興機構

データ共有のメリット

- データの早期公開はよりよい成果が期待できる
 - エラーの早期発見、早いコミュニティ形成
- 一つのデータから多様な研究
- 再現可能性
- 他データとの結合
- 学際的研究の促進
- データの保全
- サイテーション
- 教育やアウトリーチ
- 社会や市民科学とのつながり

データ共有のデメリット

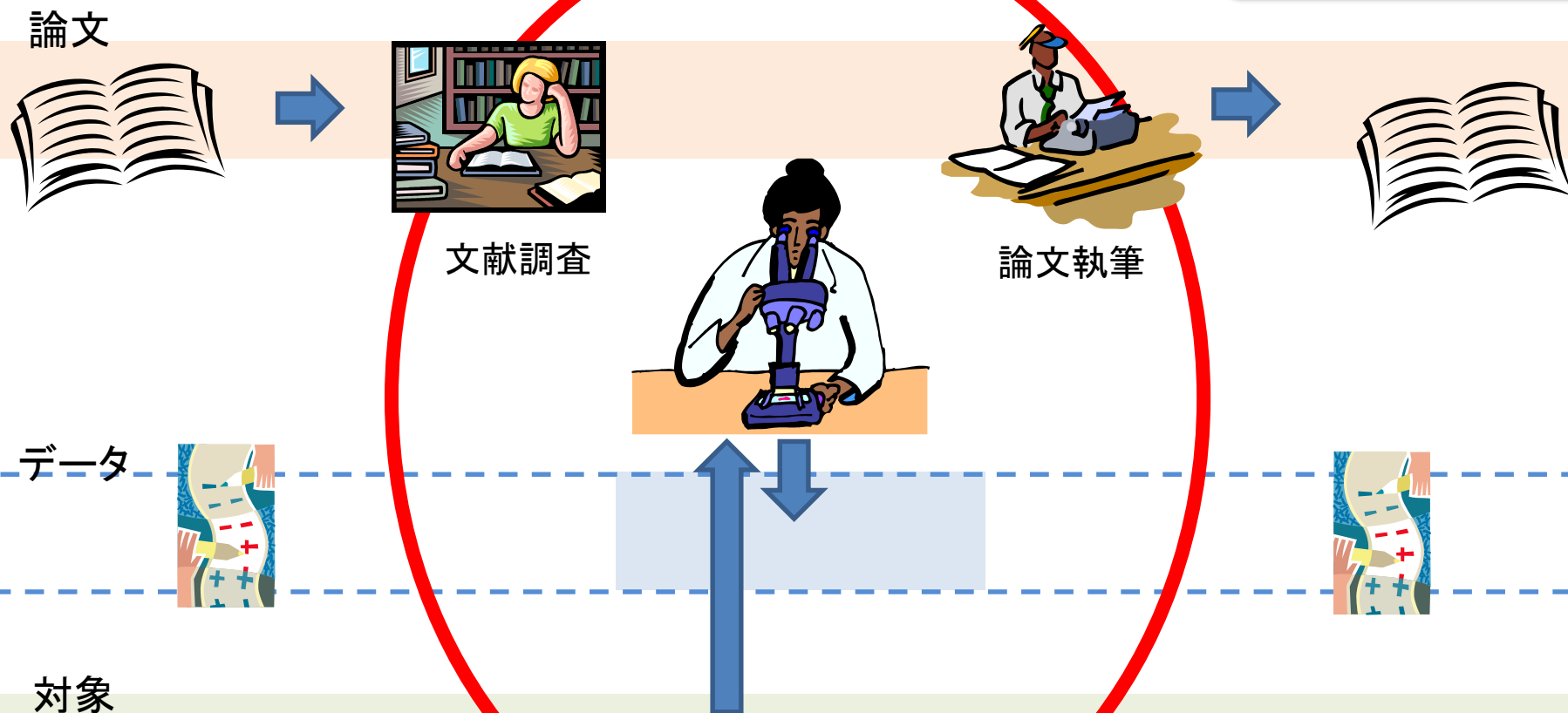
- 内部利用より高度な“標準化”の必要
- キュレーション
- 維持コスト
- 横取り研究の可能性

Internet/Web時代の研究

- すべてがデジタルへ

デジタル化以前の研究者

研究と執筆



現在の研究者

研究と執筆とデータ生成

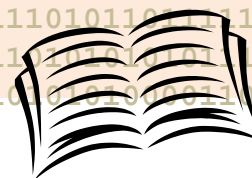
論文



文献調査



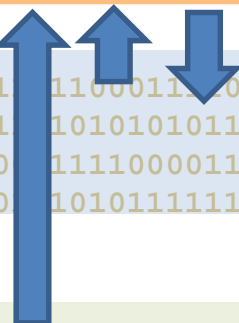
論文執筆



データ



データ利用



データ公開



対象

今後の研究者

研究成果＝データ生成

論文・データの一体化

論文



データ利用

データ公開

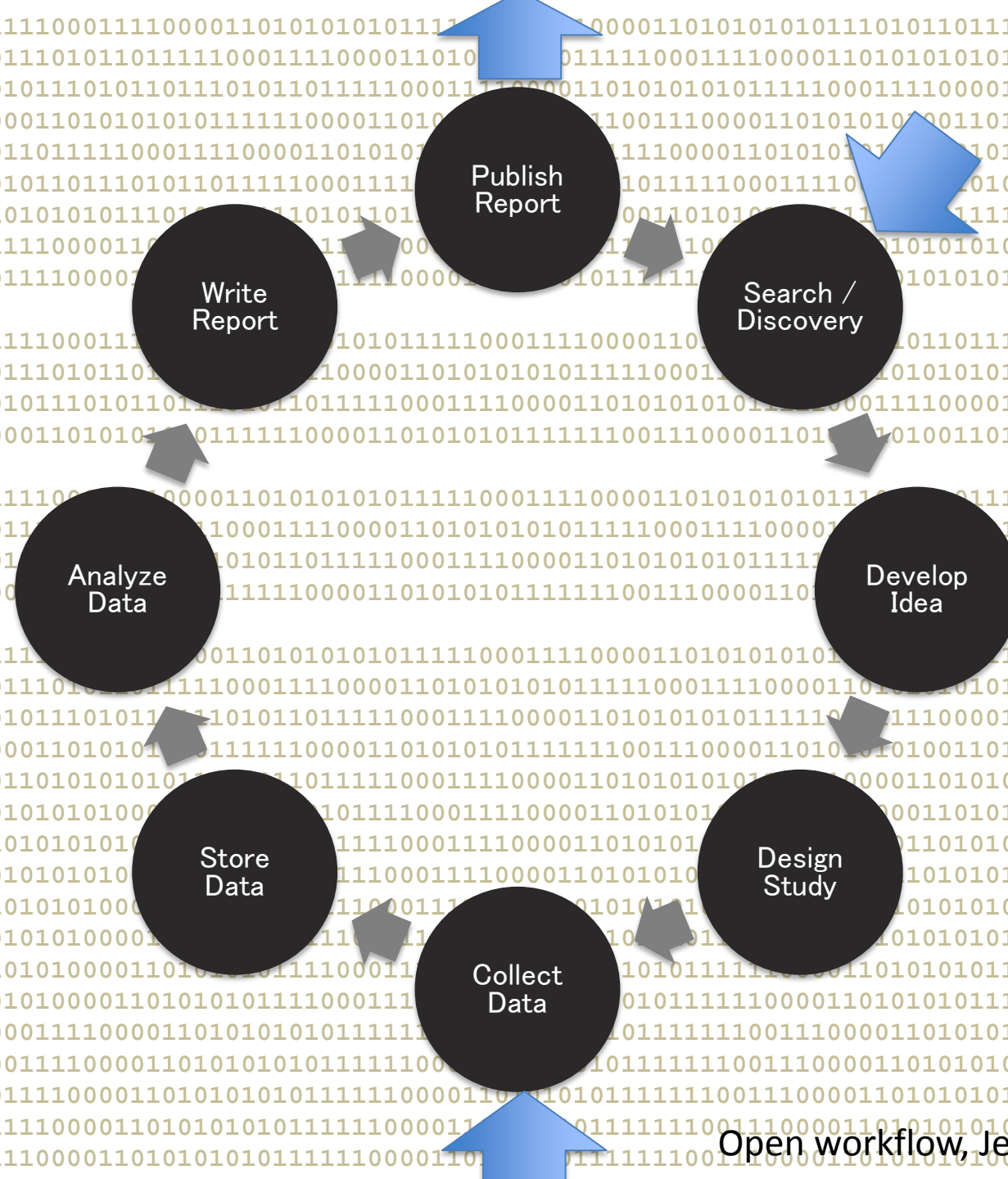


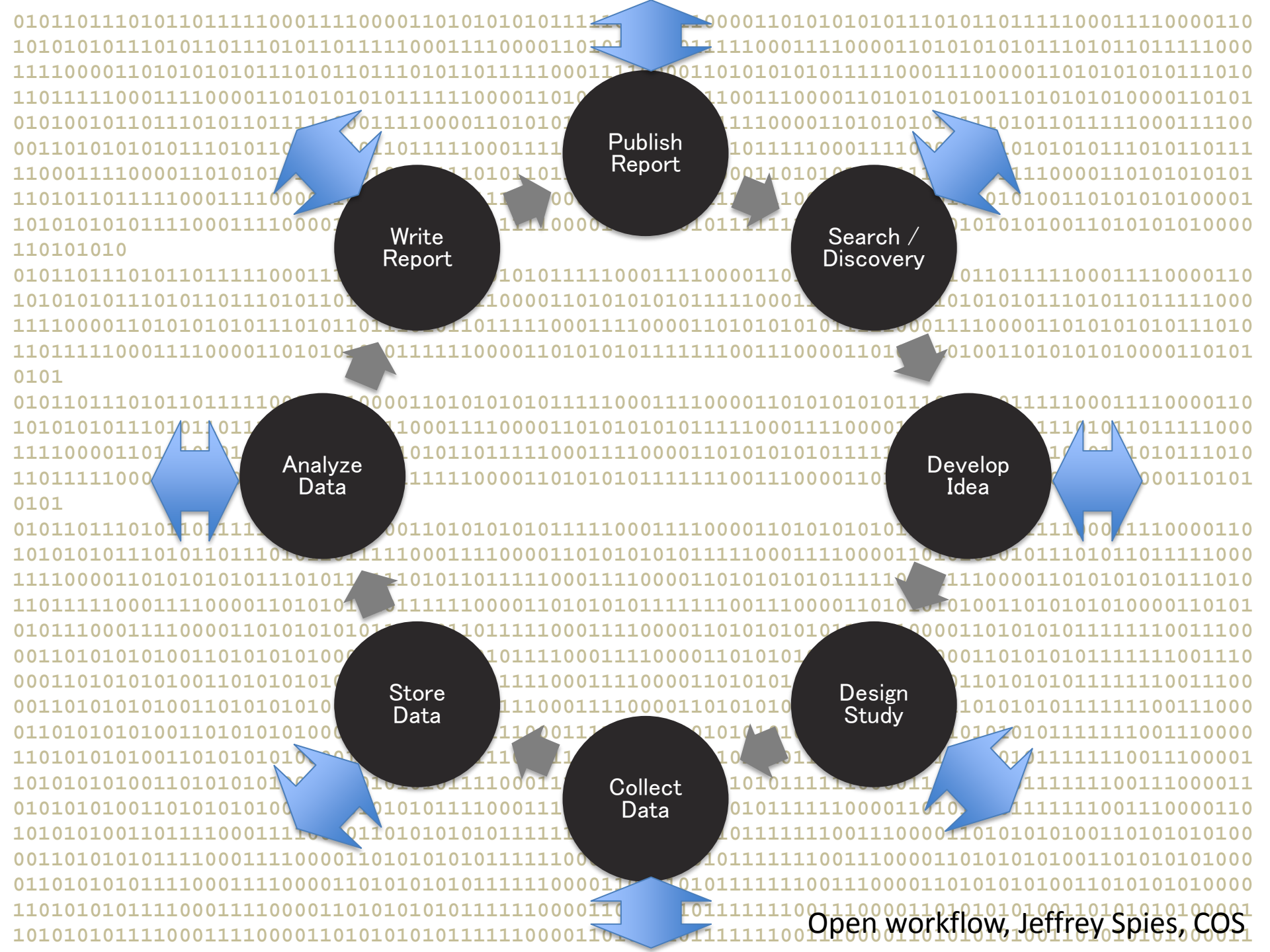
データ



対象



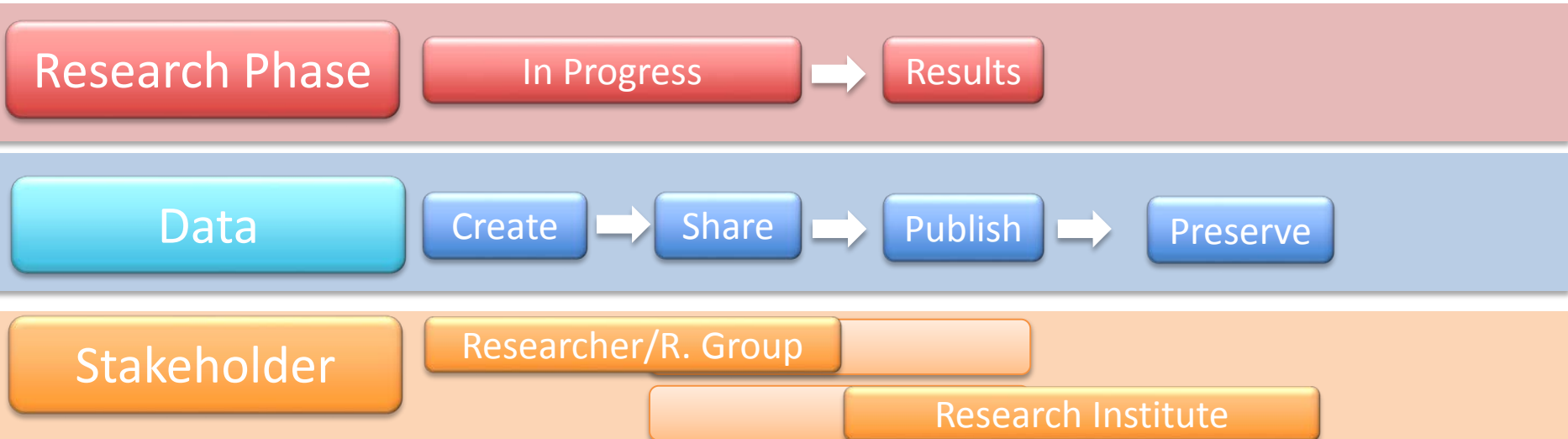




Open workflow, Jeffrey Spies, COS

Data Life Cycle

- データは作られ, **共有され**, **公開され**, 保存される
 - 多くは共有からオープンに公開へ
 - 一部(プラバシーデータ、セキュリティデータ)は共有のまま
- データのライフサイクルを通じたサポート



FAIR原則

● Findable 見つけられる

- (メタ)データはグローバルで永続的でユニークな識別子を持つべき
- データは豊富なメタデータで記述されるべき
- (メタ)データは検索可能な資源に登録あるいはインデックス化されるべき

● Accessible アクセスできる

- (メタ)データは標準的な通信プロトコルで識別子を使って取得できるべき
- プロトコルはオープンでフリーで汎用の実装可能であるべき
- プロトコルは必要であれば認証、認可の手順を持つべき
- メタデータはデータが入手不可になってもアクセス可能であるべき

● Interoperable 相互運用可能

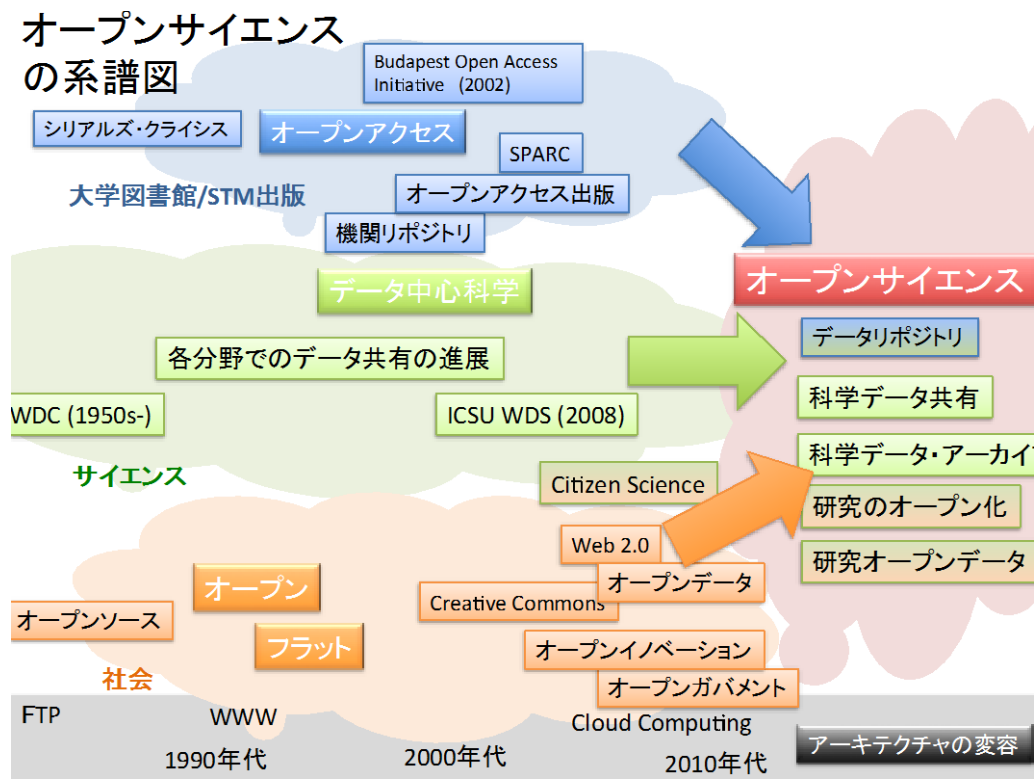
- (メタ)データは知識表現として形式的かつアクセス可能かつ共有可能かつ広く適用可能な言語を使うべき
- (メタ)データはFAIR原則に沿った語彙を使うべき
- (メタ)データは他の(メタ)データへの適切な参照を持つべき

● Re-usable 再利用できる

- メタ(データ)は精度と関連性に関する属性を複数持つべき
- (メタ)データは明確でアクセス可能なデータ利用ライセンスを付与すべき
- (メタ)データは由来をつけるべき
- (メタ)データは領域に関連したコミュニティの標準に合致すべき

研究データ共有に関わる活動

- Research Data Alliance (RDA) 2013-
 - 多様なステークホルダーの集まり
 - 研究者
 - 大学・研究機関
 - 出版社
 - ファウンダー
 - 政府関係
 - ITベンダー
 - 企業



RDA

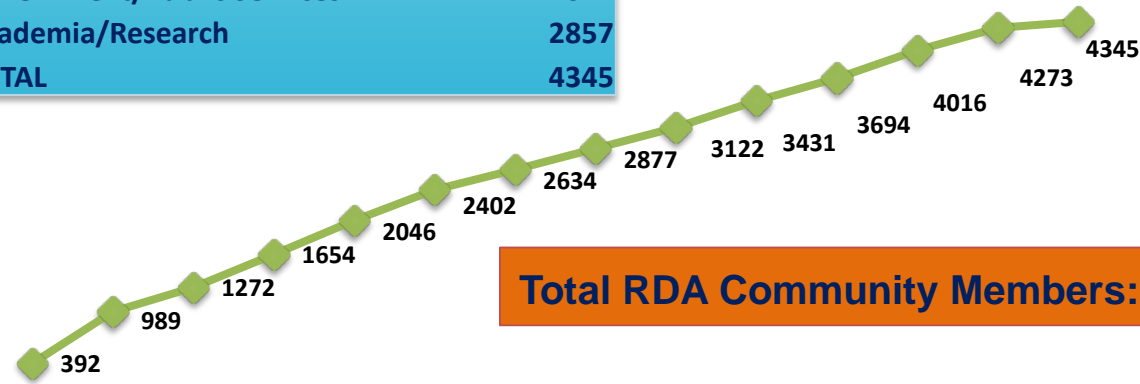
- RDA is an international organization focused on the development of infrastructure and community activities that reduce barriers to data sharing and exchange, and the acceleration of data driven innovation worldwide.
- RDAはデータ共有と交換の障害を減らすための**インフラストラクチャー**とコミュニティ活動の発展と世界的なデータドリブンイノベーションの加速に焦点を当てた国際的な活動である。
- RDA is building the social and technical bridges that enable open sharing of data to achieve its vision of researchers and innovators openly sharing data across technologies, disciplines, and countries to address the grand challenges of society.
- RDAは、研究者と**イノベーター**が技術・ディシプリン・国境を越えてオープンにデータを共有するというビジョンを達成するために、データのオープンな共有を可能とする**社会的**・技術的架け橋を作る。

RDAの活動

- 幾つかのトピックス
 - 再現性
 - データ保存
 - 領域リポジトリのベストプラクティス
 - カリキュラム開発
 - データサイテーション
 - データタイプレジストリ
 - メタデータ
 - ...

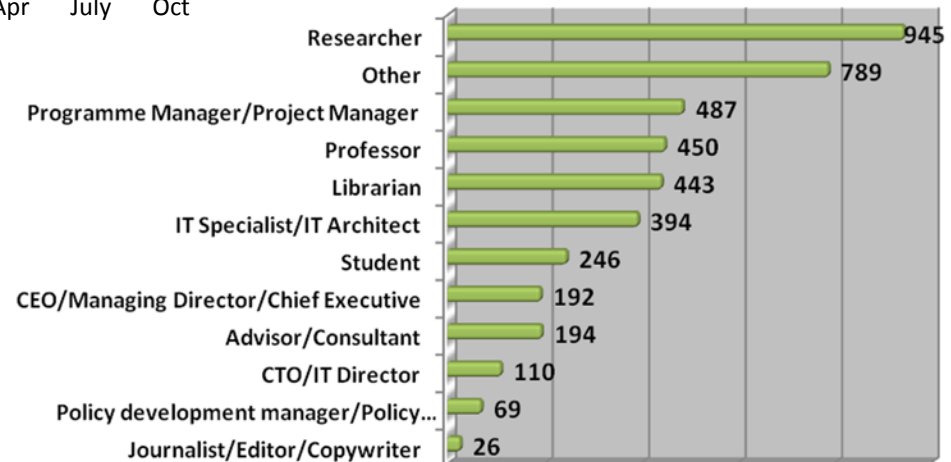
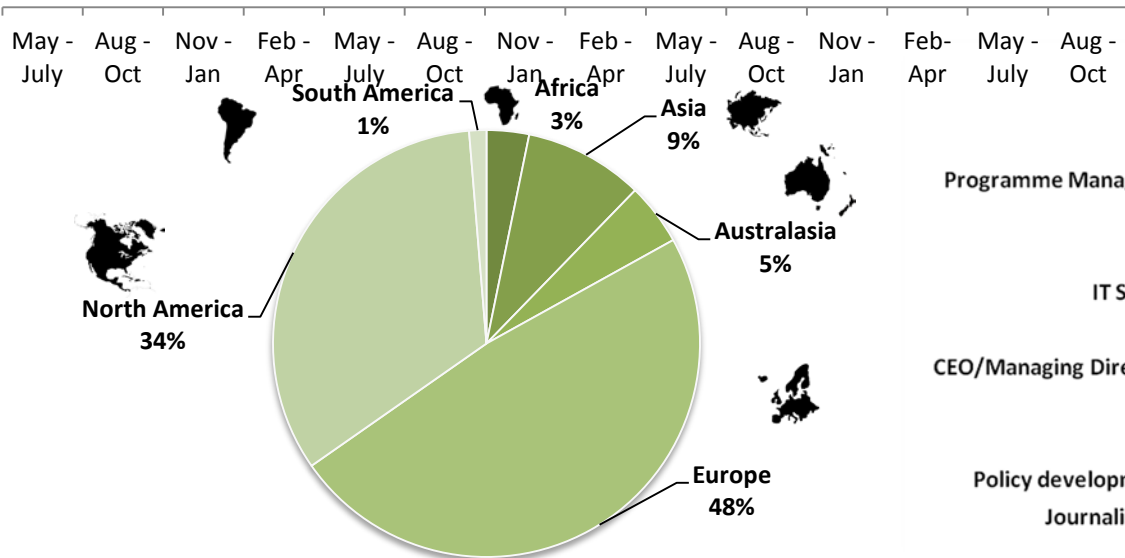
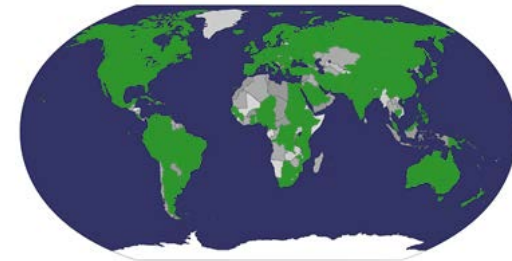
Type	Members (Sept. 2016)
Press & Media	27
Policy/Funding Agency	64
Large Enterprise	99
IT Consultancy/Development	143
Small and Medium Enterprise	249
Other	235
Government/Public Services	671
Academia/Research	2857
TOTAL	4345

Who is RDA?



Total RDA Community Members: 4345

from 111 countries





Organisational & Affiliate Members

44 RDA Organisational Members



6 RDA Affiliate Members



German Data Forum



研究データ利活用協議会

Research Data Utilization Forum (RDUF)

- 2016年6月発足
- ジャパン・リンク・センターの活動の一環として設立
- 機関会員
 - 科学技術振興機構(JST),
 - 物質・材料研究機構(NIMS),
 - 国立情報学研究所(NII),
 - 国立国会図書館(NDL),
 - 情報通信研究機構(NICT),
 - 千葉大学附属図書館/アカデミック・リンク・センター
- 個人会員

研究データ利活用協議会

Research Data Utilization Forum (RDUF)

- 目的

- 研究データに関する多様なセクターを集める
 - 実務者を中心に
- 研究データの共有と公開に関する課題を共有する
 - 分野を超えて共通の課題
 - 技術の共有
- 研究データの共有と公開に対しての、技術的・社会的解決に関する議論を行う
- 海外など関連する組織・活動との情報共有とコラボレーションを図る

研究データ利活用協議会

Research Data Utilization Forum (RDUF)

● 活動計画

－ 研究会（年3回程度）

- キックオフミーティング(7/25)
- 第1回 (10/3)→
- 第2回 (10/26) ここ
- 第3回 人文科学データ？

－ 公開シンポジウム等

- サイエンスアゴラ内シンポジウム

－ 研究データの利活用の未来

- » オープンサイエンスの実現手段
- » 11/4 (金) 13:30-15:00

ジャパンリンクセンター(JaLC)
研究データ利活用協議会 第1回研究会



● 研究データ共有による
イノベーションの創出
● 第8回RDA総会等の国際議論を踏まえて

2016年10月3日(月) 14:00～17:30 (開場:13:30)

第8回研究データ同盟 (RDA) 総会は、本年9月15～17日に米国デンバーで開催される研究データに関する国際会議です。会議参加者の発表を中心に、研究データを取り巻く国際情勢にスポットをあて、最新動向の共有を図ります。ディスカッションでは、研究者、図書館員をはじめ様々な立場の人々が、研究データの共有によるイノベーションの創出について共に考える場にしたいと考えています。

- ◆ 講演「RDAの概要とJaLC研究データ利活用協議会について」(仮)
国立情報学研究所教授、協議会会長 武田 英明
- ◆ 講演「オープンサイエンスを巡る世界の最新動向」
情報通信研究機構統合ビッグデータ研究センター研究統括、協議会副会長 村山 泰啓
- ◆ 第8回RDA総会参加報告
国立情報学研究所特任准教授 戴川 圭
国立情報学研究所助教 込川 悠介
科学技術振興機構知識基盤情報部長 小賀坂 康志
物質・材料研究機構技術開発・共用部門科学情報プラットフォームエンジニア 田辺 浩介
国立国会図書館電子情報部電子情報企画課 山口 聡
- ◆ Digital Infrastructures for Research 2016参加報告
国立情報学研究所准教授 山地 一嶺
- ◆ 講演「研究データと国の科学技術政策の観点からRDA総会を見る」(仮)
文部科学省科学技術・学術政策研究所 科学技術予測センター上席研究官 林 和弘
- ◆ フロアも交えたディスカッション

GENERATING INNOVATION BY RESEARCH DATA SHARING BASED ON THE INTERNATIONAL DISCUSSIONS

- 会場** 国立国会図書館 東京本館 新館講堂
東京都千代田区永田町1-10-1 東京メトロ永田町駅より徒歩10分
- 申込方法** 国立国会図書館ホームページ「イベント・展示会情報」の当該イベントページ (<http://www.ndl.go.jp/jp/event/events/201610rda.html>) より申込みを受付

参加無料
定員200名
(先着順)

<お問い合わせ先>
国立国会図書館 電子情報部 電子情報企画課 連携協力係
電話: 03-3506-5239 E-mail: jalc@ndl.go.jp





研究データ利活用協議会
サイエンスアゴラ内 公開シンポジウムのご案内

研究データの利活用の未来
- オープンサイエンスの実現手段 -

本年6月に設立された「研究データ利活用協議会」の紹介や、洪水予想やダム操作などの身近なデータ活用事例を題材に、研究データの利活用の未来を、来場者の皆さまと共に語りあげます。

日時： 2016年11月4日(金) 13:30~15:00 (受付開始 13:00~)

会場： サイエンスアゴラ内
A会場 (日本科学未来館) 7階会議室2

定員： 90名

参加費： 無料

申込： 参加を希望される方は
こちらからお申し込みください。

https://japanlinkcenter.org/top/index.html#top_event



プログラム

時間	内容	講演者
13:30-13:45 (15分)	「研究データ利活用協議会」の紹介	武田 英明【本協議会会長】 (国立情報学研究所 教授)
13:45-14:15 (30分)	「水災害分野におけるデータ活用最前線 - 洪水予測からダム操作、都市浸水予測まで -」	洪尾欣弘 (土木研究所水災害・リスクマネジメント国際センター 専門研究員 政策研究大学院大学 連携准教授)
14:15-14:35 (20分)	対談	洪尾欣弘、武田英明
14:35-15:00 (25分)	フロアも交えたディスカッション	モデレータ 武田 英明【本協議会会長】 (国立情報学研究所 教授)

お問い合わせ先 Japan Link Center (JaLC) 事務局
国立研究開発法人 科学技術振興機構 知識基盤情報部
メールアドレス info@japanlinkcenter.org

(参考) サイエンスアゴラについて

サイエンスアゴラは「あらゆる人に開かれた科学と社会をつなぐ広場」の総称です。サイエンスアゴラ2016は、11月

まとめ

- オープンサイエンス
 - Internet/Webの発展とともに
 - 技術、習慣、マーケット、ルールのそれぞれが変化
- オープンサイエンスの重要なステップとしての研究データ共有
 - データ公開の原則(FAIR)
 - データから研究ワークフロー全体へ
- 横断的な対話が始まっている
 - Research Data Alliance (RDA)
 - 研究データ利活用協議会(RDUF)