

2015.3.9 第4回 SPARC Japan セミナー2014

「グリーンコンテンツの拡大のために我々は何にをすべきか？」

於 国立情報学研究所 12階 1208,1210会議室

図書館によるデータ管理 への道筋

情報・システム研究機構 国立極地研究所

南山 泰之

minamiyama@nipr.ac.jp

目次

1. 本セミナーの位置づけ
2. なぜデータが重要なのか？
3. 海外の動向
4. 国内での取り組み
5. 図書館が果たすべき役割の検討

「グリーンコンテンツ」の再定義

- そもそも”green contents”とは？
 - green journalに掲載された「論文」を機関リポジトリに搭載したもの
 - 研究データや博物館資料、紀要などは対象(想定?)外
- だが・・・
 - Green(研究機関) vs Gold(出版者)
 - 若葉が青々と育つイメージ = “コンテンツの拡大”を連想

そこで、本セミナーでは

「学術研究機関が機関リポジトリ等において公開・発信する学術コンテンツ」

として再定義。

オープンアクセス

“無料”かつ“誰でも利用できる”
点を強調

- オープンアクセスの定義は・・・

- “Budapest Open Access Initiative” (2002年2月)

- 「・・・インターネット上において無料で利用可能であり・・・合法的目的のための利用が・・・財政的、法的また技術的障壁なしに、誰にでも許可されること」
- 「・・・著作権が持つ唯一の役割は、著者に対して、その著作の同一性保持に対するコントロールと、寄与の事実への承認と引用とが正当になされる権利とを与えること・・・」

“著作者の明示”、“正当な評価と引用”
のみを要求

オープンアクセスの広がり

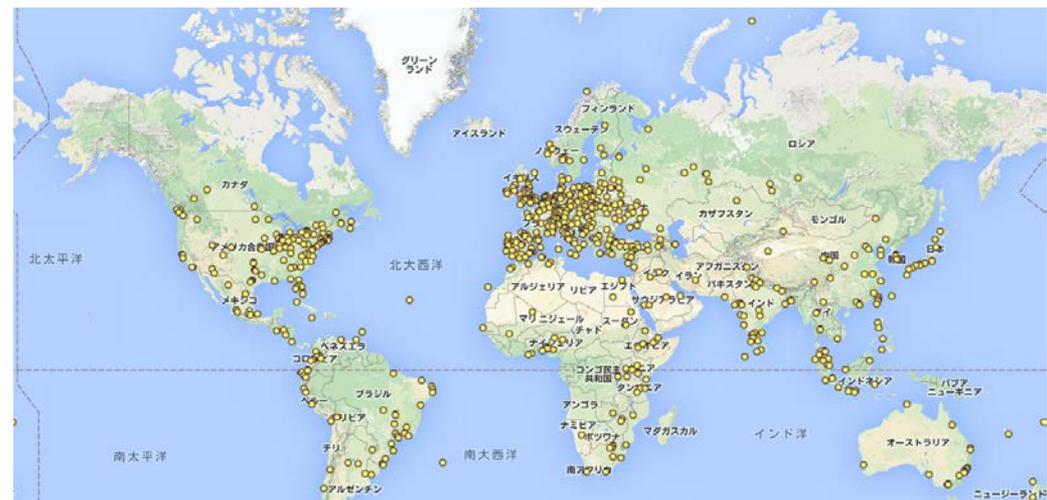


Open Access map

<http://www.openaccessmap.org/>

Content Types

- [Repository](#)
- [Journal](#)
- [Policy](#)
- [Other](#)



オープン”アクセス”から オープン”サイエンス”へ

- オープンサイエンスとは
 - ICTの発達により生まれたアプローチ
 - 新たな知の創造プロセス
 - 従来の論文のオープン化に留まらない
 - 根拠となるデータ、研究プロセスの公開
- この概念の広がりは・・・
 - 異なる分野へのイノベーションの引き金に
 - 利活用の促進、公共への還元

本セミナーの位置づけ

- オープンサイエンスに期待されること
 - 知識の循環、再利用によるイノベーション
 - ⇒ 他業界との連携促進
- ✓ つまり、データの連携、管理手法の検討を目的とする本セミナーは、オープンサイエンス活動の一環。
- ◆ では、なぜ「データ」が重要なのか？

「データ」とは？

- 博物館における「データ」
 - － 資料画像データ
 - － メタデータ
- 研究・観測データ
 - － 数値データ
 - － テキストレコード
 - － イメージ・ビジュアルデータ
 - － メタデータ



言語	日本語
統一資料名 (大分類)	金沢大学旧蔵教育掛図
発行日	2013-10-01
時代についての情報	受入：明治38（1905）年3月28日
トピック	掛図
数量データ	本紙寸法：106.4×66.9cm 1
注記	印：「第四高等学校図書」、「明治卅八年三月廿八日第四高等学校図書室」 ラベル：「11門62類328号」 注記：「Section showing Elevation-crater theory」、「Diagram-section of a normal Volcano」 収蔵場所：書庫、箱⑤ 状態：八双継目少し破れ、吊下注意、本紙ややシミ・汚れ
由来	第四高等学校旧蔵
形態	紙、軸装、手描

✓ 様々な研究の根拠・成果

Data Summary

Baseline Meteorological Data in Siberia (BMDS) Version 5.0

◀ Back View Meta Data Record ▶▶ Download ▶▶ Quick Look ▶▶ Document ▶▶

Select Version 5.00 ▼

External link

http://www.jamstec.go.jp/acd/ag/Summary.action?selectFile=A20110307_001&downloadList=ast-top

Summary

This dataset is the baseline meteorological data in Siberia that was constructed and quality-controlled for the GAME-Siberia Project and RIGG/JAMSTEC. It consists of daily data of main meteorological elements for 107 stations in Lena River basin and other territories of Siberia for 1950 - 2006.

DATA CONTRIBUTOR & GROUP

- Hironori Yabuki (Research Institute for Global Change/Japan Agency for Marine-Earth Science and Technology)
- Hotaeek Park (Research Institute for Global Change/Japan Agency for Marine-Earth Science and Technology)
- Rikie Suzuki (Research Institute for Global Change/Japan Agency for Marine-Earth Science and Technology)
- Tetsuo Ohata (Research Institute for Global Change/Japan Agency for Marine-Earth Science and Technology)
- Haruko Kawamoto (Research Institute for Global Change/Japan Agency for Marine-Earth Science and Technology)

PARAMETERS

- Air Temperature
- Humidity
- Surface Pressure
- Shortwave Radiation
- Longwave Radiation
- Net Radiation
- Precipitation Amount

Spatial coverage

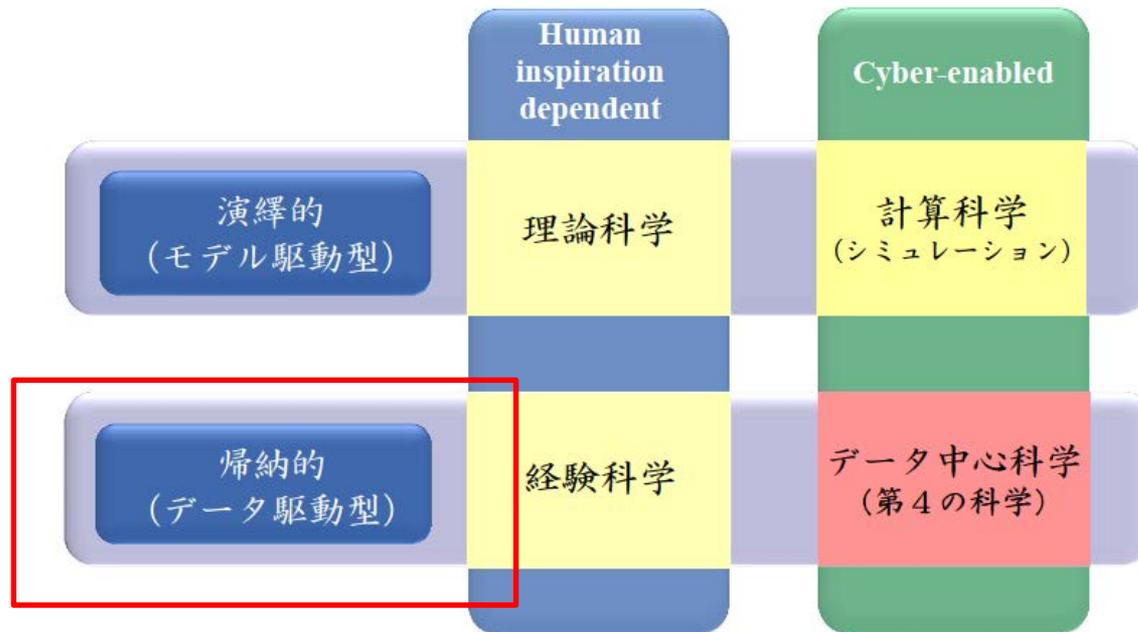
N: 73.5° S: 48.52° E: 180.0° W: 55.1°

Temporal coverage

1950-01-01 - 2006-12-31

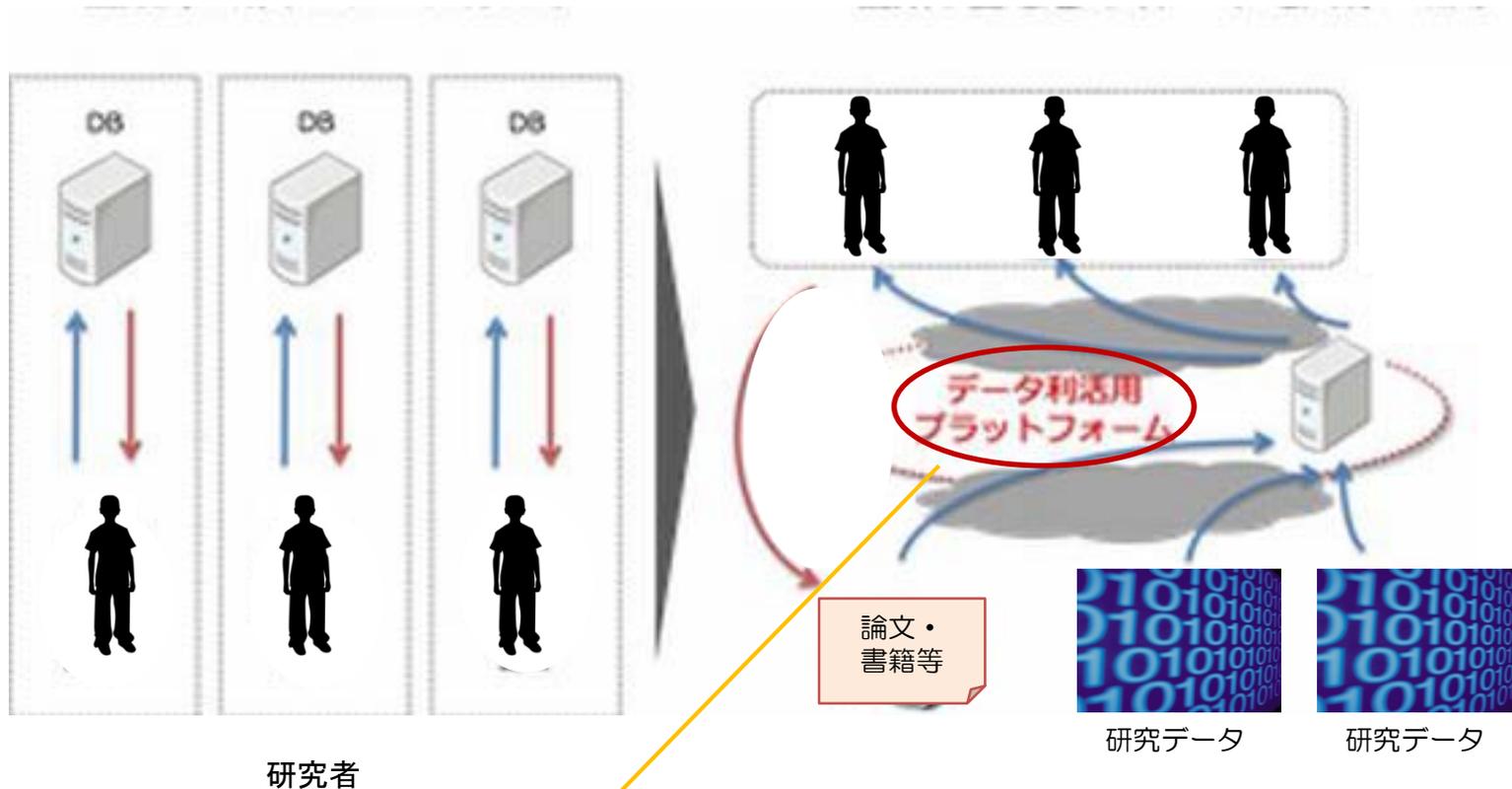
アカデミック・ビッグデータの 効果的な開発と利用

データ中心科学（第4の科学）



ビッグデータが拓く新しい科学研究の在り方.
<http://www.ism.ac.jp/~nobuo/BIGDATA/20130527-Kitagawa.pdf>

データ駆動型の研究



保存・公開を担う
"プラットフォーム"の重要性

経済産業省HP(一部改変):
<http://www.meti.go.jp/press/2014/06/20140609004/20140609004.html>

一方では

「研究データガバナンス」の問題

あなたの研究室は大丈夫ですか？

秘密漏洩事故や不正疑惑で訴訟の当事者となる管理責任者→



2~3年後



論文は公開されても、元データは散逸↑

「研究データガバナンス」の知的基盤強化は、研究不正を防止し疑惑から研究者を守る

© 豊田哲郎 理化学研究所 生命情報基盤研究部門 平成21年11月13日 委員会用資料

<http://database.riken.jp/item/cria42s2ria42s184i/>

豊田哲郎. 理化学研究所生命情報基盤研究部門平成21年11月13日委員会用資料

<https://database.riken.jp/sw/pic/ja/cria42s2ria42s184i/>

データ保存・公開の意義

- データ保存・公開は、
 - “オープンサイエンス”の前提
 - 異なる分野へのイノベーションの引き金
 - 公共への還元
 - (共有による)研究の効率化
 - 国際的プロジェクトなど
 - 研究不正への対応
 - 研究の透明性
 - 「再現可能性」を高める

海外での活動

公開支援:



大学図書館の取り組み:



Purdue University Research Repository

PURR

参考: <http://oui-oui.jp/2014/09/rdm/>

データの業績化

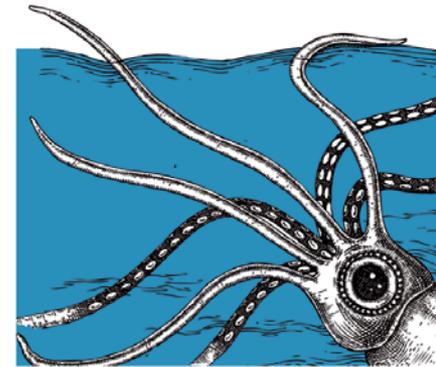
The screenshot shows the NSF website's Grant Proposal Guide (GPG) page. The header includes the NSF logo and the tagline "WHERE DISCOVERIES BEGIN". A navigation bar lists categories like FUNDING, AWARDS, DISCOVERIES, NEWS, PUBLICATIONS, STATISTICS, ABOUT NSF, and FASTLANE. The main content area is titled "Grant Proposal Guide" and "NSF 13-1 January 2013 GPG Summary of Changes". It details significant changes to the merit review criteria, including updates to Chapter II.C.1.e (Proposal Certifications) and Chapter II.C.2.b (Project Summary).

COMMENT

The reuse factor

The reference is not dead — it is exploding to encompass the full spectrum of research outputs from lines of code to video frames, explains **Mark Hahnel**.

Researchers are still struggling to find, manipulate and cite research outputs other than published papers. Data-management plans for research — detailing what data will be created and how, and outlining plans for data sharing and preservation — are a core requisite of all grant applications for a long list of US and UK funding agencies. These include the US National Science Foundation, the US National Institutes of Health, NASA, the UK Biotechnology and Biological Sciences Research Council, the UK Medical Research Council and the Wellcome Trust. Funders require grantees to make all their products available in a citable, sharable and discoverable man-



their research to have quantifiable impact. To this end, the Research Data Alliance (RDA) was established in August 2012 by a steering group of funding agencies from the United States, Europe and Australia. The RDA aims to accelerate and facilitate research data sharing and exchange across multiple disciplines that have complicated funder mandates and a need to cite various unconventional research outputs. An RDA working group plans to provide prototypes and examples. For instance, individual cells of a spreadsheet or a few frames of a video can be cited in a way that does not dilute a paper's total number of citations. Early adopters of open-data science are already seeing the benefits. Computer scientist

NSF (米国国立科学財団)

2013年1月から業績としてデータを記載可能に
“Publication” → “Products”

“GPG Summary of Changes”. NSF 13-1 January 2013

リユースファクター (提案)

データやコードの引用によるインパクトを測定

Hahnel, M. The reuse factor. *Nature*. 16 October 2013, 502(7471), p.298. doi:10.1038/502298a

転載元:

「オープンサイエンスを支えるデータライブラリアン」

<http://www.slideshare.net/ikeuchiui/2014-41488937>

国内での取り組み



大学博物館等と連携しながら、データリポジトリの普及を实践(2012?~)



研究データへのDOI付与実験プロジェクト(2014.9~)
管理組織や体制の整備

データ管理を巡る国内動向

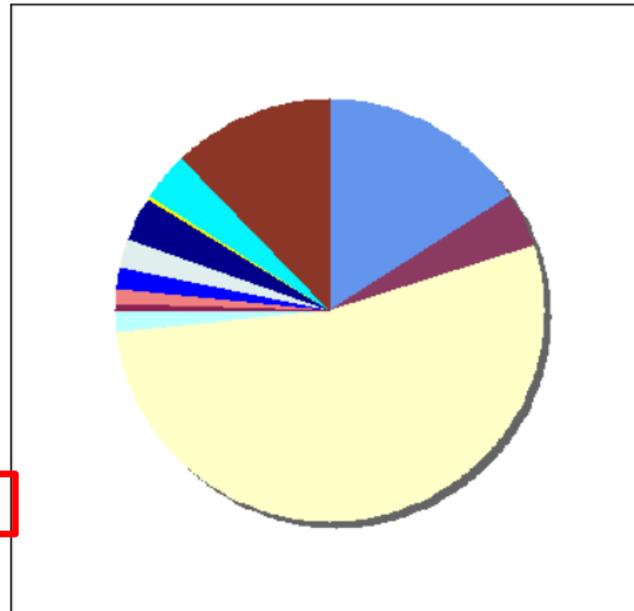
- 国際的動向を踏まえたオープンサイエンスに関する検討会
 - 内閣府、2014年12月～
- 科学技術情報整備審議会/第四期科学技術情報整備基本計画策定に向けた基本方針検討部会
 - 国立国会図書館、2014年12月？～
- その他
 - 日本学会会議学術フォーラム(2015年3月4日)
 - 「科学を変えるデータジャーナル – 科学技術データの共有・再利用の新たなプラットフォーム構築へむけて–」

で、日本の図書館は？

蚊帳の外？

機関リポジトリによる データの取り扱い現状

Journal Article(学術雑誌論文)	227,812 (15.7%)
Thesis or Dissertation(学位論文)	63,757 (4.4%)
Departmental Bulletin Paper(紀要論文)	774,801 (53.4%)
Conference Paper(会議発表論文)	22,897 (1.6%)
Presentation(会議発表用資料)	5,633 (0.4%)
Book(図書)	19,649 (1.4%)
Technical Report(テクニカルレポート)	21,861 (1.5%)
Research Paper(研究報告書)	32,301 (2.2%)
Article(一般雑誌記事)	49,657 (3.4%)
Preprint(プレプリント)	373 (0.0%)
Learning Material(教材)	4,054 (0.3%)
Data or Dataset(データ・データベース)	52,420 (3.6%)
Software(ソフトウェア)	29 (0.0%)
Others(その他)	176,290 (12.1%)
合計	1,451,534



・クリックすると詳細な図が別ウインドウで表示されます

IRDBコンテンツ分析(2015年1月現在):
<http://irdb.nii.ac.jp/analysis/index.php>

機関リポジトリ推進委員会 コンテンツWGでの活動

1. 2014年8月 WG発足
2. 2014年11月 総合展での研究集会発表
3. 2015年2月 国際WGとの共同による
IDCC (International Digital Curation
Conference) 参加
1. 2015年3月(予定) 報告書作成

研究データの動向把握及び試行収集 班のタスク

1. なぜ図書館が研究データを扱うのか？を明らかに
2. (主に海外の)研究データ管理の動向調査
3. 学術雑誌論文の機関リポジトリ登録に併せた、付随データの公開勧誘方法の検討
(2014年12月～)

なぜ図書館が研究データを扱うのか？を明らかに

- 研究データ保存・公開の意義
 - オープンサイエンスの前提
 - 共有による研究の効率化
 - 研究不正への対応
- なぜ大学図書館が研究データを扱うのか？
 - 大学図書館本来の役割に合致
 - 大学設置基準第38条
 - データリポジトリのニーズ
 - 資料の組織化・機関リポジトリ運営の経験
 - 専門家集団による、保存の持続性担保

参考：日本における研究データリポジトリへの課題・調査経過報告

<http://id.nii.ac.jp/1280/00000020/>

(主に海外の)研究データ管理の動向 調査

- 機関のポリシー
 - 助成機関から出されているテンプレートの試訳
 - 海外大学の事例
- メタデータ
 - 研究データのメタデータ標準とJuNii2の比較
 - DataCite、DIF、etc...
 - JaLC2の動向
- システム面
 - 適切なプラットフォームの検討
 - DSpace ? CKAN ?
 - JAIRO Cloudの可能性

海外機関からの情報収集

- 2014年11月4～6日
 - 英国エジンバラ大学図書館、Digital Curation Centreとの情報交換
- 2015年2月9～12日
 - International Digital Curation Conferenceへの参加・発表
- 2015年2月13～16日
 - サウサンプトン大学、エクセター大学との情報交換

(私見) 実際のところ、海外での実務 はどんな感じ？

- ポリシーの策定 ≡ 蔵書構築方針
 - リポジトリに入れられるデータの種類を決める
 - 収集対象かどうかの判断を誰がするのか？
- 普段の仕事は、**メタデータの付与が中心**
 - 電子資料相手の目録業務？
 - 対象とするデータの分野によって、メタデータ項目が変わる？
 - 入力のための仕様書、マニュアルの整備が重要。
- 利用のためのガイダンスなど

論文付随データの公開勧誘方法の検討

- WG委員所属大学等（千葉大、岡山大、極地研）のコンテンツ収集活動にて検討中。
 - 学術雑誌論文のリポジトリ搭載依頼時にメール勧誘を行う
 - 千葉大学にて試行開始（2015年2月～）
 - 紀要論文の投稿受付時、個別に勧誘を行う
 - 極地研にて試行開始（2015年2月～）

推進委員会で検討すべきこと

✓ ポリシー

- 機関におけるデータ管理ポリシー
- Funding Agencyとの相互理解、情報共有

✓ システム

- 研究ワークフローを視野に入れたシステム開発
- 研究成果の先取権や透明性を担保

✓ 人材育成

- 専門職員の育成
- 図書館員とURAとの連携

2015年3月現在までのまとめとして、
オープンサイエンスに関する機関リポジトリの現状と展望：
http://www8.cao.go.jp/cstp/sonota/openscience/3kai/3_openscience_shiryu_1_3.pdf

アプローチの違い

- 研究者は・・・
 - － 最初にデータありき
 - － 博物資料、観測・研究データを元に、論文やデータを作成
- 図書館員は・・・
 - － 最初に論文ありき
 - － 論文の根拠としてのデータのみを意識しがち
 - ⇒ データの一部しか捕捉できず、研究者には使いにくい
- ✓ 林立するリポジトリ、散らばるデータ
 - ⇒ 相互運用性の不存在

先行事例は、意外と近くにある

- 現状は・・・
 - 論文 → 図書館
 - 研究データ → 研究者(コミュニティ)
 - 博物資料 → **博物館** + 研究者(コミュニティ)
- 図書館の外に目を向ければ、モノのデータ化、リポジトリでの取り扱いの事例は存在する
 - 研究データも博物資料も、最初にモノありき
 - 研究フローに沿った運用？

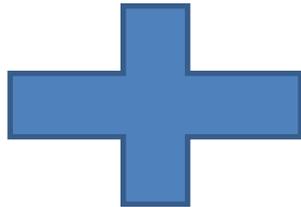
機関リポジトリの役割を振り返る

- “...a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members.”
 - 大学とその構成員が創造したデジタル資料の管理や発信を行うために、大学がそのコミュニティの構成員に提供する一連のサービス

Clifford A. Lynch. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age" 2003

将来的な目標

- ✓ 機関リポジトリの定義も踏まえて、
 - 論文 → 図書館
 - 研究データ → 図書館 + 研究者 (コミュニティ)
 - 博物資料 → 博物館 + 研究者 (コミュニティ)



- ✓ 博物資料等も含めた、横断的な情報管理体制モデルの構築