

平成24年度京都大学図書館機構講演会
2012.12.18 京都大学附属図書館

新たな書誌フレームワークと 統合的発見環境

佐藤 義則

東北学院大学文学部

yssat@tscc.tohoku-gakuin.ac.jp

目次

1. 目録の対象と目的の変化

- 資料のデジタル化の再考

2. 利用の変化

- 「SCREAL調査2011」の結果から

3. 新たな発見環境

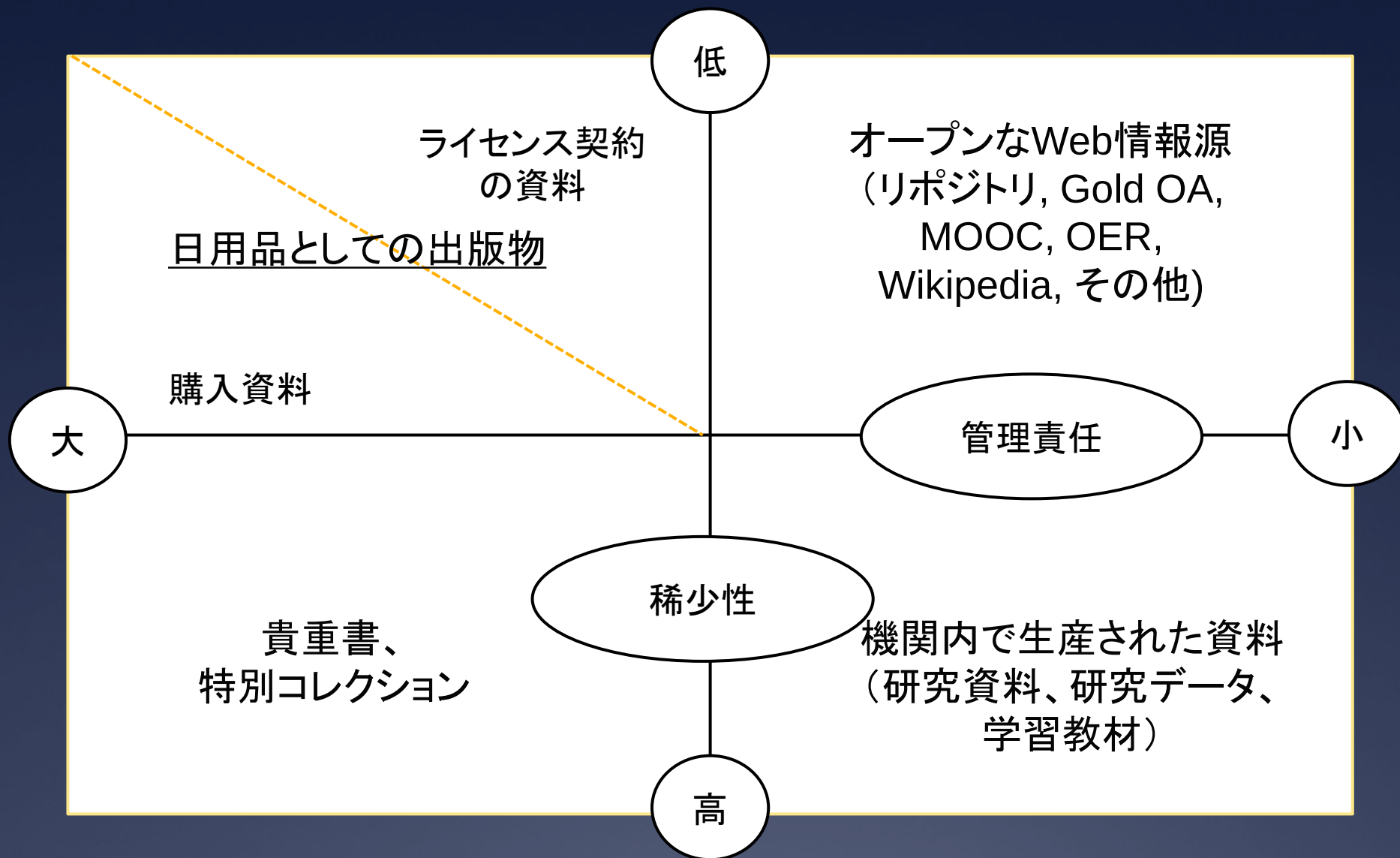
- 新たな書誌フレームワークへ向けた展開
- LODがもたらす可能性とは？

4. これから何が必要か

- 新たな連携協力の必要性和基盤形成に向けて

1. 目録の対象と目的の変化

- かつて： 館内資料の検索手段として
 - 研究において、図書館に資料があることが不可欠な時代
 - ファインディング・ツール or 現物の代替物としての目録
- これまで： 資源共有の基盤として
 - 外部機関をも対象にした検索のサポート
 - 資料の存在の確認と相互利用
 - ✓ 基盤としての、NACSIS-CAT/ILL
- これから： 資料の発見とアクセスのための基盤として
 - 異なる粒度の検索との融合
 - 所蔵ではなく、何らかの方式での「アクセス可能性」の保証が重要に



引用元: Malpas, Constance. "Scarcity and Abundance: the Cooperative Imperative in Special Collections," 53rd Annual RBMS Preconference, 20 June 2012.
available at <http://www.oclc.org/research/presentations/default.htm> (2012-07-04).



アルキメデスのパリンプセスト

出典: <http://www.archimedespalimpsest.org/>

デジタル化の進展

- 電子ジャーナル
 - OAジャーナル、OAメガジャーナル、カスケード出版、、
- eBooks
 - 欧米の出版社プラットフォーム
 - 出版デジタル機構
 - Kindle, iPad, kobo, e-Book reader,,
- マス・デジタイゼーション: Google Books, Internet Archive – Open Library, HathiTrust, DPLA,,, 国立国会図書館

Google Books Settlement

2005.9- 10	The Authors GuildとAAP(Association of American Publishers)がGoogle を著作権侵害で提訴（クラスアクションとして）
2008.10.28	訴訟当事者3社が和解を申立て
2009.9.18	米国政府、裁判所に対し旧和解案の成立に反対する意見書提出
2009.10.7	裁判所、訴訟当事者に対し和解条項の修正を命令 （Googleはこの時点までに書籍1000万冊以上をデジタル化）
2009.11.13	訴訟当事者、裁判所に対し修正和解条項提出（修正和解案）
2009.11.19	裁判所、修正和解案を予備承認
2010.2.4	米国政府、裁判所に対し修正和解案の成立に反対する意見書提出
2010.2.18	裁判所、修正和解案に関する公聴会を開催 （Googleはこの時点までに書籍1200万冊以上をデジタル化）
2011.3.22	裁判所、修正和解案を認めない裁定を下す

国立国会図書館

- 2009年度補正予算：127億円
- スキャンのみ（デジタル化テキストは作成しない）、館内公開のみ；ただし、許諾の得られたものについてはネット上で公開
- 著作権法の改正（2010.1施行）
- 国立国会図書館サーチ（2012.1 - ）

所蔵資料のデジタル化の対象範囲

資料群	年代等	デジタル化資料提供場所
古典籍	貴重書・準貴重書、江戸期以前の和漢書等	http://dl.ndl.go.jp/#classic
図書	明治期以降、1968年までに受け入れた図書	http://dl.ndl.go.jp/#books
雑誌	明治期以降、2000年までに発行された雑誌	http://dl.ndl.go.jp/#magazines
官報	1883(明治16)年7月2日(創刊)～1952(昭和27)年4月30日に発行された官報	http://dl.ndl.go.jp/#kanpo
博士論文	1991～2000年度に送付を受けた論文	http://dl.ndl.go.jp/#etd

デジタル化資料提供状況（平成24年11月末時点）

資料種別	デジタル化資料提供数(概数)		
	インターネット公開	館内限定提供	合計
古典籍	7万点	2万点	9万点
和図書	32万点	57万点	89万点
和雑誌	0.5万点	102万点	102.5万点
官報	2万点	-	2万点
博士論文	1.5万点	12.5万点	14万点
憲政資料	0.01万点	-	0.01万点
合計	43万点	173.5万点	216.5万点

国立国会図書館「資料デジタル化について」

<http://www.ndl.go.jp/jp/aboutus/digitization.html>

新たなタイプの 電子的情報資源共有

- HathiTrust
<http://www.hathitrust.org/>
- Internet Archive “Open Library”
<http://openlibrary.org/>
- DPLA (Digital Public Library of America)
<http://cyber.law.harvard.edu/research/dpla/>

HathiTrust

- 共同の研究用コレクションの保存とアクセス提供
 - ✓ Google Books、Internet Archive由来 + 自前のコンテンツ
- 約1,100万の電子化資料(3割弱がパブリックドメイン)
- 訴訟問題(Authors Guild等による提訴)
 - ✓ フェアユース
- 『研究用コレクションのクラウド・ソーシング:大規模電子化後の図書館環境における印刷体の管理』※
 - ✓ 利用頻度の低い図書の管理をHathiTrustのようなデジタル・リポジトリおよび共有の印刷体保存リポジトリへ外部委託することの可能性と得られる効果(図書館スペースの節約、コスト削減)
 - ✓ “大学図書館がHathiTrustと連携し、大規模電子コレクションへのパブリックアクセスの拡大を進めること”(2011.1)

※Malpas, Constance. *Cloud-sourcing Research Collections: Managing Print in the Mass-digitized Library Environment*. OCLC Research, 2011.1, 76 p.
<http://www.oclc.org/research/publications/library/2011/2011-01.pdf>

電子情報資源の確保と共有コレクション構築

- JISC

- 「コンテンツと電子化・プログラム」(2004 – 2009)

- 「eコンテンツプログラム」(2009 -)

- 「コンテンツプログラム」(2011 -)

A. 電子化とオープン教育情報源(9プロジェクト)、B. 大規模電子化(7プロジェクト)、C. 電子コンテンツのクラスタリング(9プロジェクト)

- JISC Collections

- eコレクションズ (2011 -)

- ✓ 商業出版社や商業プロバイダ等による電子資料の調達と提供

- ✓ 有料制(機関による支払い)

コレクション構築の変化

- 紙媒体でのコレクション構築
 - 資料は「所蔵」
 - 利用者が図書館にやって来ることを前提
 - 事前の選書と受入(＋整理)が重要
 - ✓ 資料の検索(特定と配置場所への案内)
- 電子情報資源：利用者(利用)中心のサービス構築
 - 資料は「ネットワーク上」
 - 利用者も「ネットワーク上」; 利用者の必要に合わせた資料の調達も可能 → PDA (Patron Driven Acquisition)
 - ✓ PDA: 図書館のコレクションに対する根本的な問題提起(？)
 - 資料のグローバルな発見可能性こそが重要

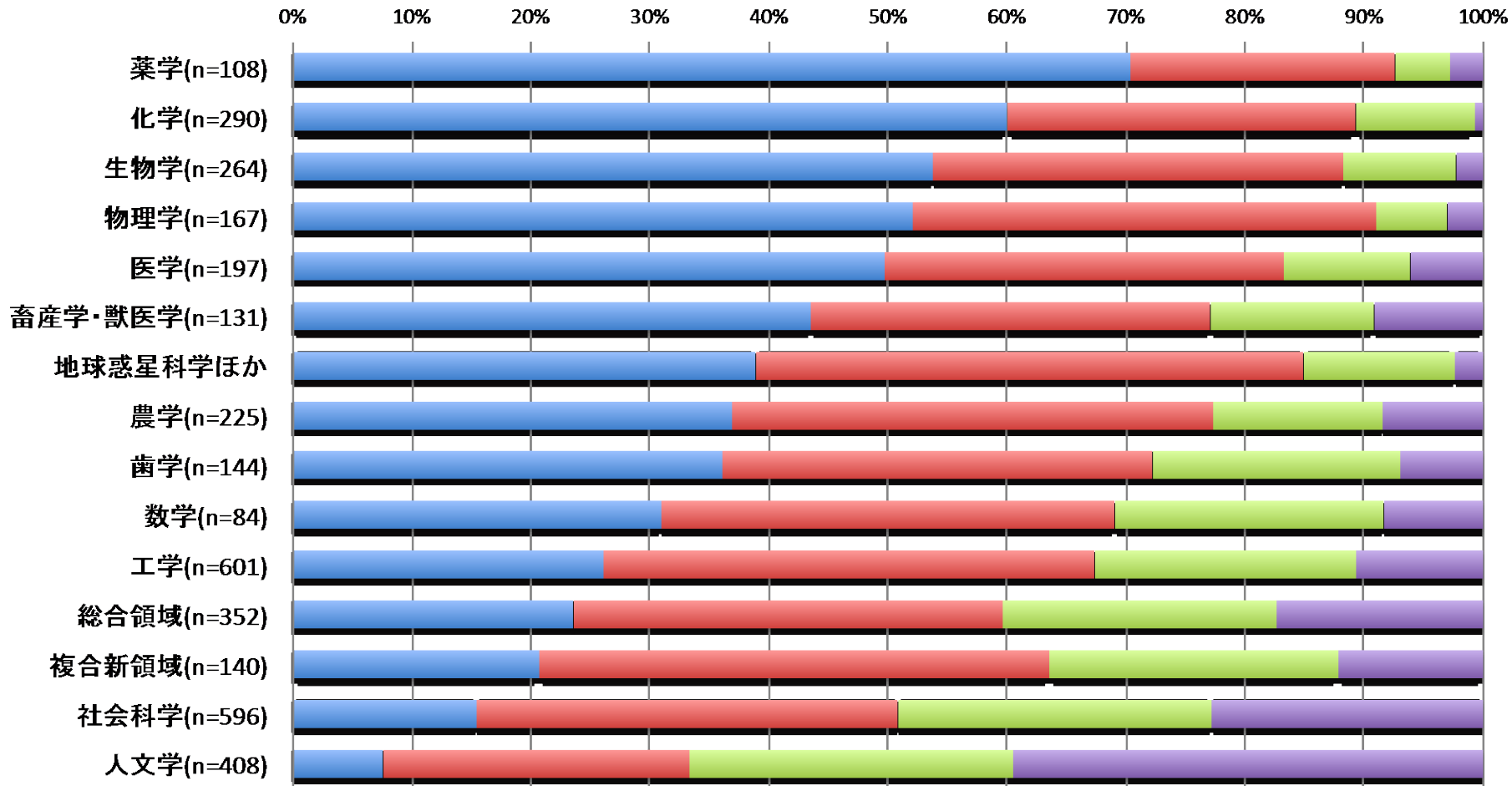
	印刷体資料	電子情報資源
資料の在処	図書館内	図書館(内)外
利用対象	(多数ある)コピーの一つ	単一(唯一)の情報源
アクセス	物理的所蔵に基づく	契約や協定に基づく
作成方式	人手による確認、入力	(プログラム等による) 既存データの有効活用
目録処理	共同分担目録 (書誌データと資源の共有)	集中的作業 (典拠データ、リンク形成)
課題	データの品質レベル	データの品質レベル 永続的アクセスの管理 情報源間の関係性の整理

2. 利用の変化

- SCREAL調査2011 <http://screal.jp/>
 - 調査対象機関： 45機関(国立大学 21、公私立大学 15、国立研究所 9)
 - 調査期間： 2011年10月12日～12月31日
 - 調査方式： Webアンケート(メールによる依頼→参加申込→回答)
 - 回答数： 3,922(完全回答数;推定回答率: 6.04%)

電子ジャーナル利用の浸透度

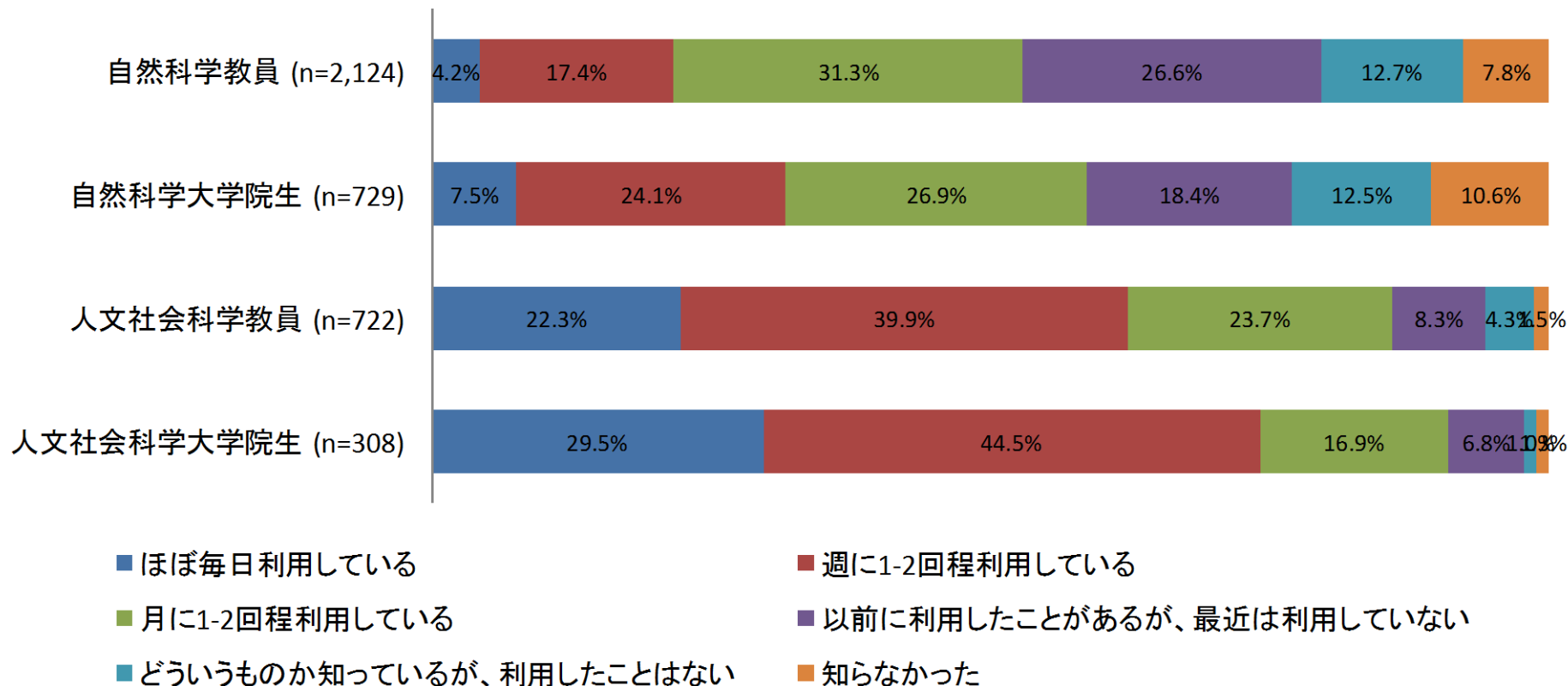
電子ジャーナルをどの程度利用していますか



■ ほぼ毎日利用している ■ 週に1-2回程利用している ■ 月に1-2回程利用している ■ 最近は利用していない、利用したことがない等

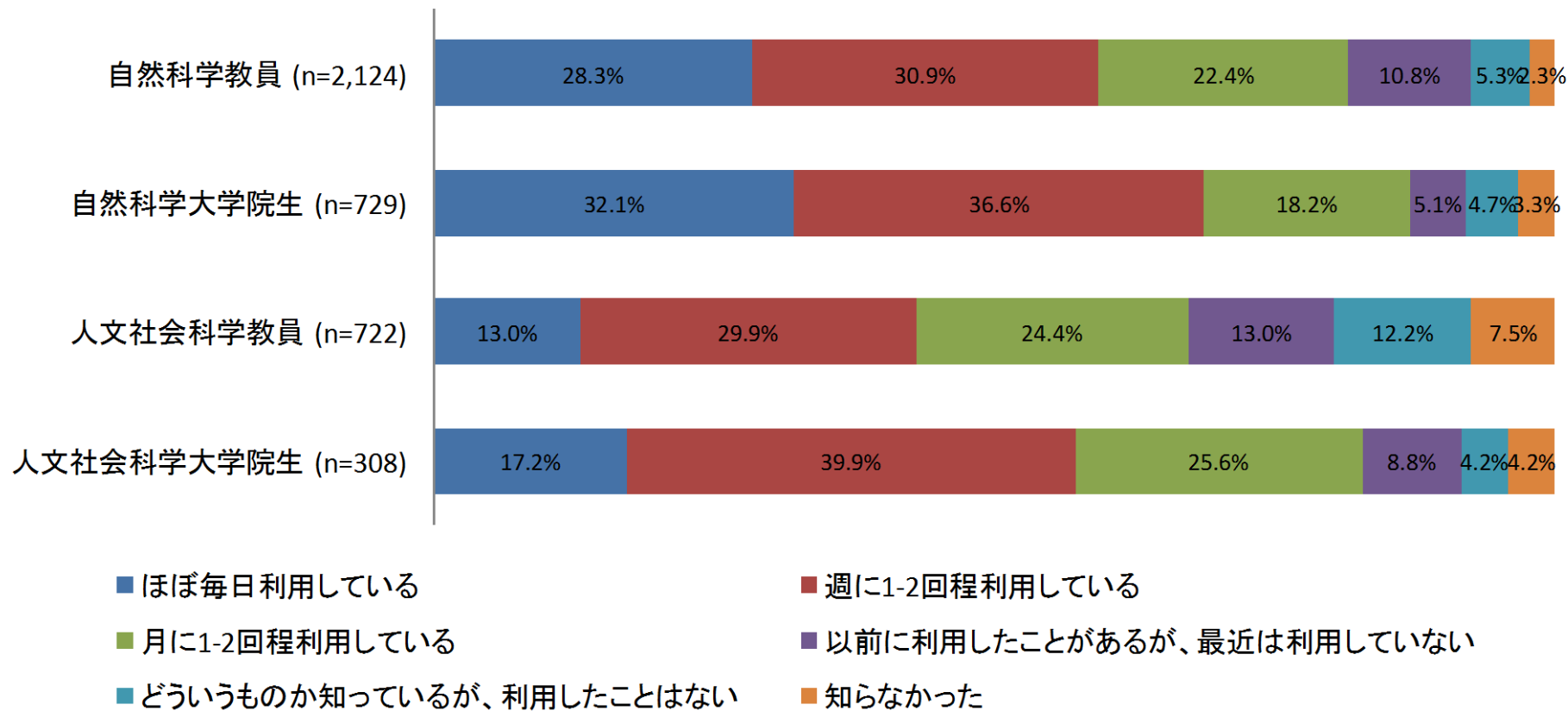
発見手段の利用頻度(Q20)

図書館のOPACやNACSIS-WebCatで検索



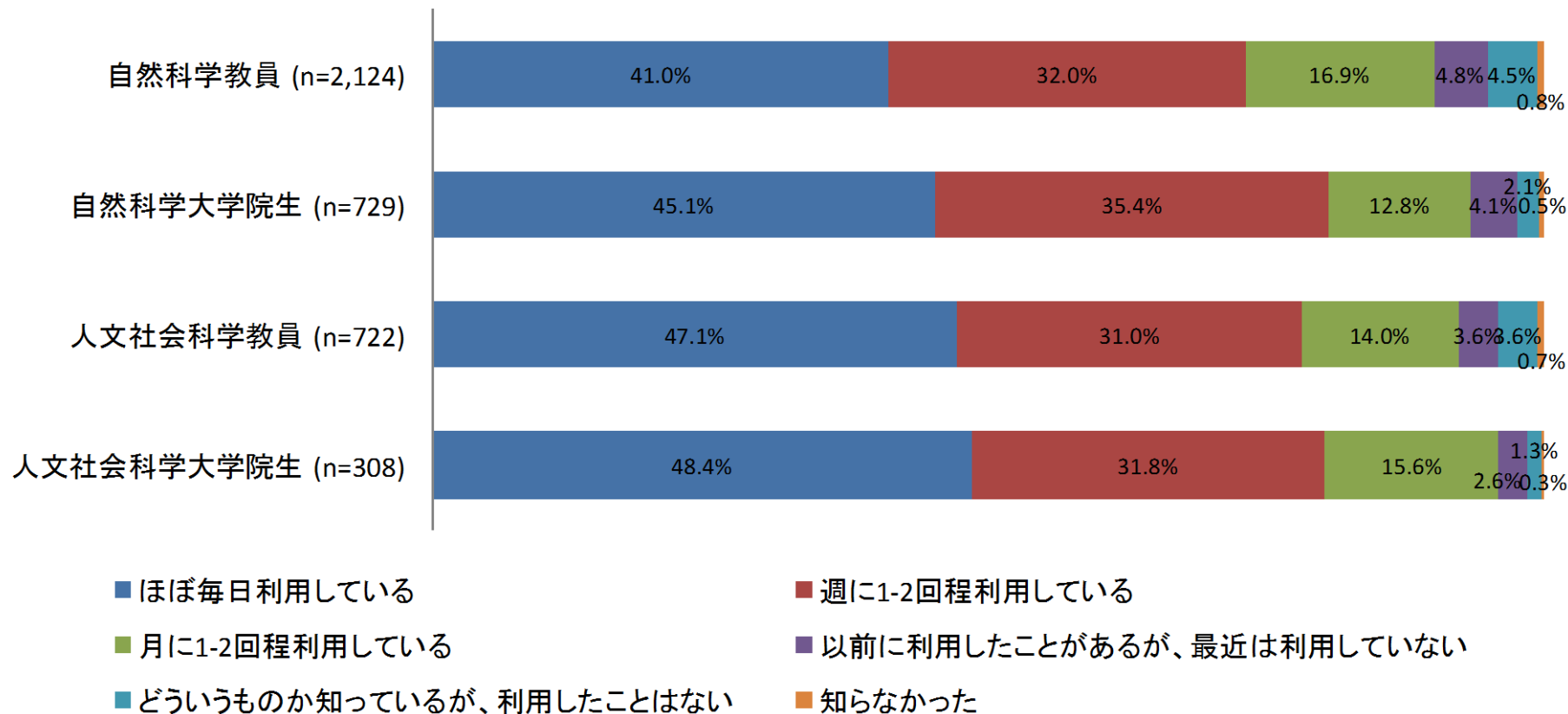
発見手段の利用頻度(Q20)

索引・抄録データベースで検索



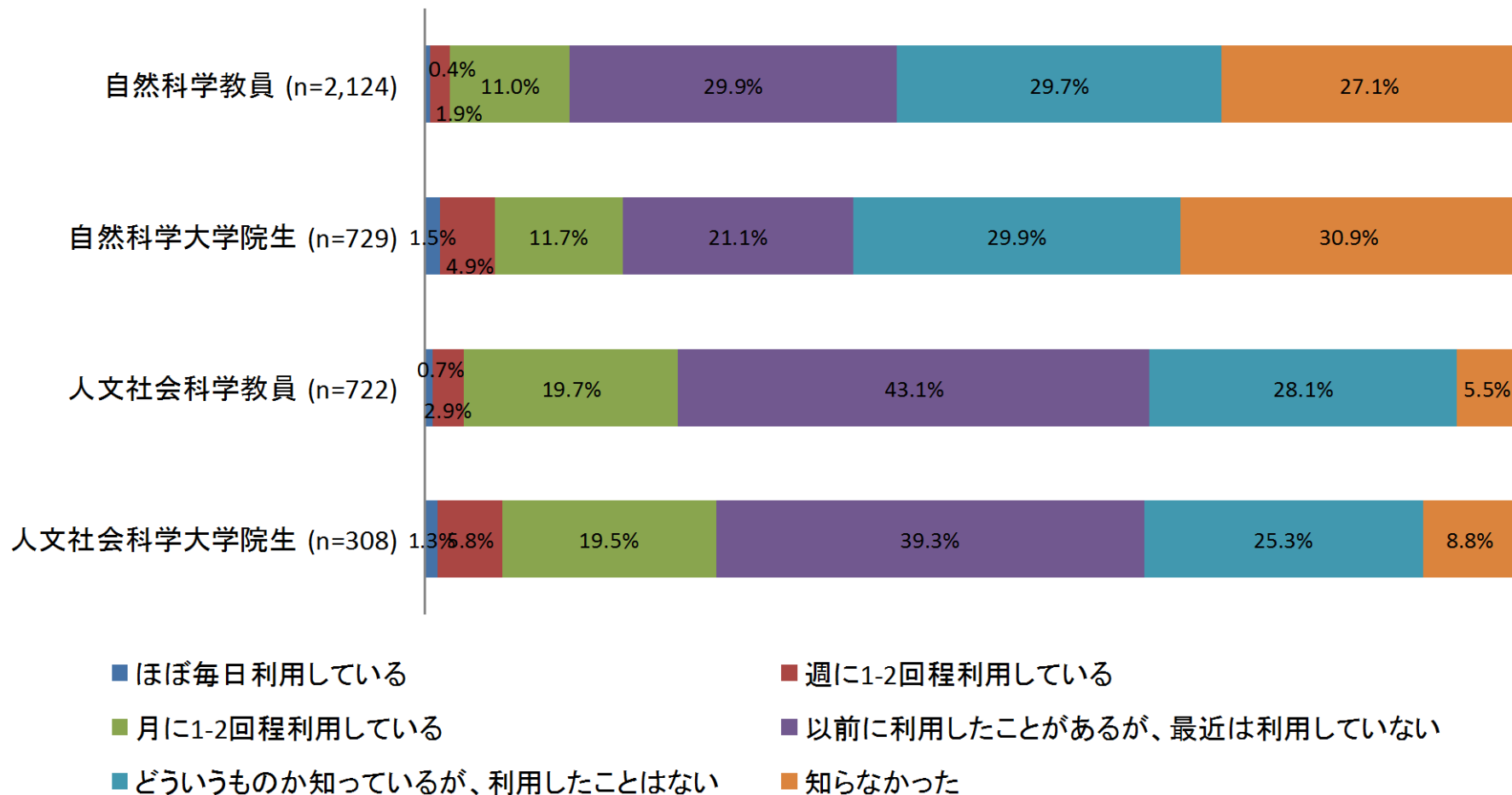
発見手段の利用頻度(Q20)

ウェブの検索エンジンで検索



発見手段の利用頻度(Q20)

図書館のレファレンスサービスを利用



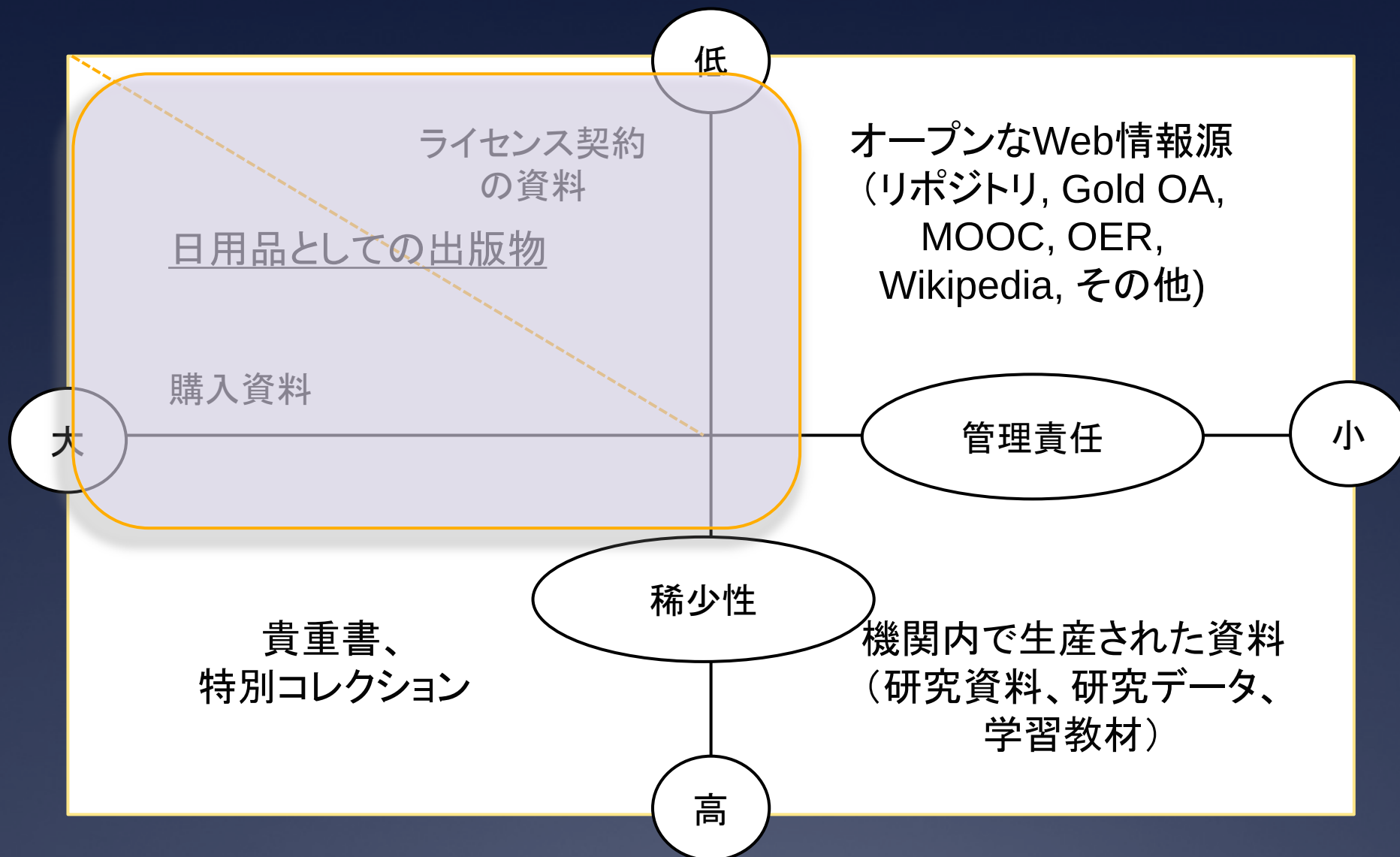
「...研究者は、ますますこれらの活動をオンライン上で行うのであるから、当然、研究図書館のサービスはデジタルの研究環境に不可欠な部分となる必要がある。実際に、大学図書館、研究図書館はもうすぐそうなることを見込むべきである。」

「良いサービスとは、研究者が研究のすべての段階で必要とするデジタル情報を発見し、利用する能力によって規定されるだろう。」

Palmer, Carole; Tefteau, Lauren C., & Pirmannet Carrie M. *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development*. OCLC Research, 2009.1, p. 34

3. 新たな発見環境

- ウェブスケールの発見環境
 - Summon, EBSCO Discovery Service, WorldCat Local,,,
 - OCLC WorldShare Management Service
- 図書館界を超えたメタデータ利用
 - 発見のための多様なプラットフォーム (Google, Amazon, Open Library,,,)
 - 流通過程におけるメタデータ利用 (ONIX, Amazon)



引用元: Malpas, Constance. "Scarcity and Abundance: the Cooperative Imperative in Special Collections," 53rd Annual RBMS Preconference, 20 June 2012.
available at <http://www.oclc.org/research/presentations/default.htm> (2012-07-04).

「情報カオス： 四つの要因」

1. 組織毎の、数多くの情報サイロ： それぞれ別個の、検索／発見の仕組み、メタデータの「システム」、そして不完全なコンテンツ記述
2. 図書館／電子図書館の発見の仕組み： 低い精度、不適切な再現率
3. 図書館が作成／収集するメタデータ： ウェブの世界と遠くかけ離れたメタデータ
4. Google等の検索エンジン： 使いやすい検索や発見の方法を提示。しかし、多くの情報資源がインデックス化の対象外であることは非周知

Keller, Michael A. “Linked data: A way out of the information chaos and toward the Sematic Web,” EDUCAUSE Review, 2011.7.<http://www.educause.edu/ero/article/linked-data-way-out-information-chaos-and-toward-semantic-web>

Linked Open Data (LOD)の展開

- W3C (World Wide Web Consortium) Library Linked Data Incubator Group 最終報告(2011.10)
- LC^{*}, BL^{*}, Europeana^{*}, Cambridge U.^{*}, Harvard U.^{*}, OCLC^{*},,,
- New York Times^{*}, BBC^{*}, Nature^{*},,,
- DBpedia^{*}, GeoNames^{*},,,
- NDL Web Authorities^{*}
- CiNii^{*}
- OCLC – WorldCatのデータを[Schema.org](http://schema.org)形式で公開(加えて、Dewey, VIAF and FAST headingsへのリンク) ODC-BY
 - ✓ ダウンロード用ファイル(250以上の所蔵館を持つ書誌データ、120万件=8,000万トリプル)

Linked Dataの4原則 (Tim Berners-Lee)

- ① ものごとの名前としてURIを使うこと
- ② 人間やコンピュータ・ソフトウェアがものごとの名前を参照したり、調べたりできるようにHTTP URIを使うこと
- ③ URIを見に行ったとき、RDFやSPARQL(XML)のような標準方式によってそれに対する有用な情報を提供できるようにすること
- ④ より多くのものごとを発見できるように、データの中に他のURIへのリンクをいれること

RDF (Resource Description Framework)

RDF (Resource Description Framework)

[p.5,6,12]

W3C (World Wide Web Consortium) が標準化しているメタデータの規格。コンピュータが理解可能な意味表現のための形式で、主語、述語、目的語の3つ組(トリプル)で情報の関係を表現する。主語はURIか空白、述語はURI、目的語はURI、文字列(リテラル)、空白のいずれかで表現する。

たとえば、「シロクマ」というタイトルは以下のように表せる。



「電子的学術情報資源を中心とする新たな基盤構築に向けた構想」

http://www.nii.ac.jp/content/archive/pdf/content_report_h23_with_glossary.pdf

LOD化のメリット

- データ品質の向上と作業の効率化： 著作、場所、人物、出来事、主題、その他に対する識別子の利用による、信頼できる情報源からの補足データとのリンク形成によって、あるいは図書館ではこれまで作成できなかった粒度の外部データとのリンク形成によって
- データの発見と利用可能性の向上： 図書館の目録データとDBpedia、GeoNames、BBC、New York Timesといった他の領域のサービスとのリンク形成、あるいは実験のためのデータセット、データ処理に使用されたモデルとのリンク形成によって
- 図書館のウェブ上での存在の強化： データ利用、再利用からもたらされる機関の可視性の向上によって
- 専用ソフトウェアからの解放： RDFやHTTPの活用により、より一般的なツールの利用の道が開かれることによって

(引用元：「W3C Library Linked Data Incubator Group最終報告書」)

LOD対応の期待される効果

1. 発見可能性の向上

- Schema.org
- Google 等の検索エンジンの品質向上
 - ✓ ドキュメント間の関係に着目した検索システムの提供

2. 図書館システムの変革

- MARCからの移行
- FRBR -> RDA のプラットフォームとしての活用
- 新たな連携可能性
 - ✓ 例: データ更新の際のアラーティング、双方向のデータ更新

新たな発見システム

- Open Library
- The Data Hub
- Freebase
- Archives Hub
- Google Knowledge Graph

目録システムの変革の動き

- 1995年 Dublin Core 開始
- 1998年 IFLA - FRBR (Functional Requirements for Bibliographic Records)
- 2005年 JSC - AACR2からRDA (Resource Description and Access)へ
- 2009年 IFLA - Functional Requirements for Authority Data

LCの新たな書誌フレームワーク

- 2008.1 書誌コントロールの将来に関する米国議会図書館ワーキンググループ最終報告書 “On the Record”
- 2010.7 – 2011.3 RDAの検証
 - 2011.6 検証結果と勧告 -> 2013.1以降の採用
- 2011.10 *A Bibliographic Framework for the Digital Age* --- MARC21からLODへの移行の宣言
- 2012.1 ALA Midwinter Meeting で移行計画の概要説明
- 2012.5 LODのためのモデリングおよび移行計画をZepheira社に委託
- 2012.11 “**BIBFRAME**” (*Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services*)

LCのLOD採用の影響

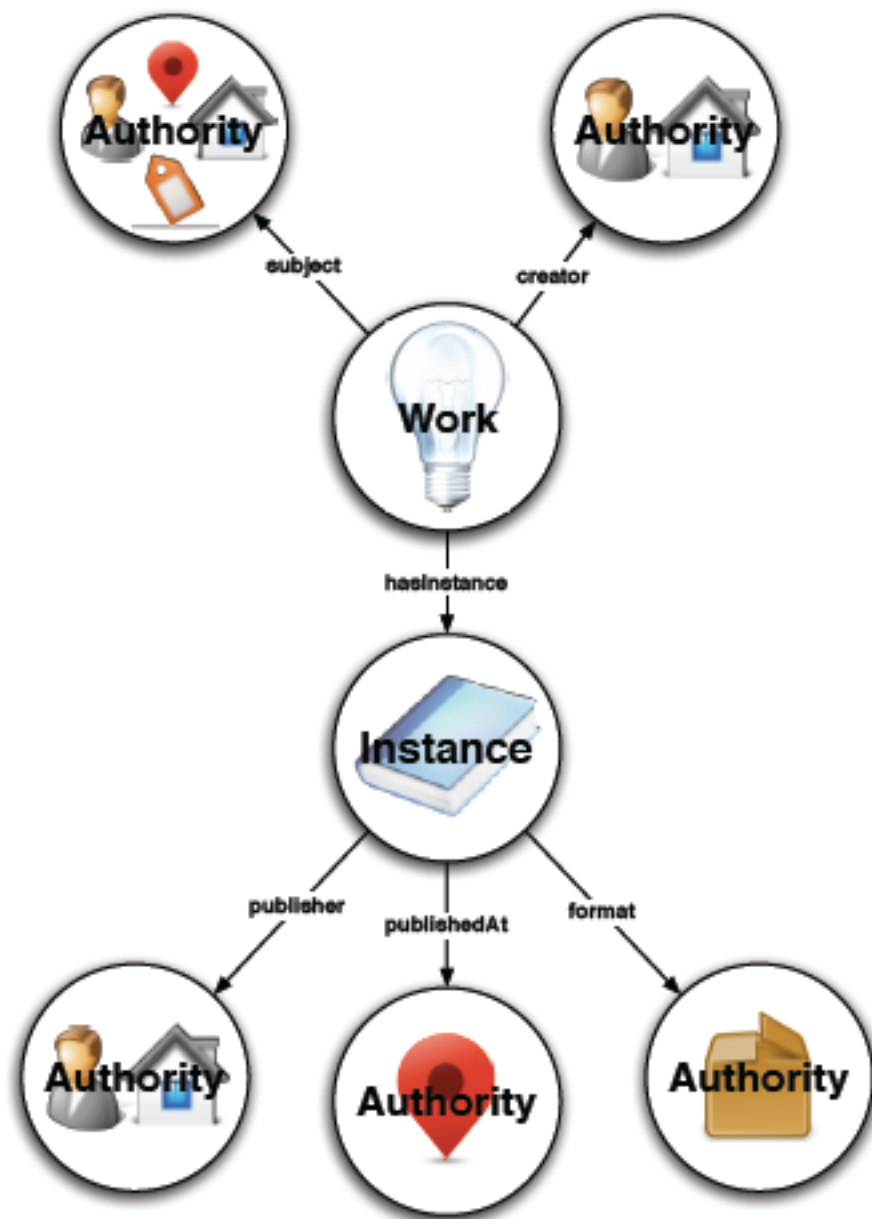
- 各国国立図書館への波及
 - 英国、カナダ、ドイツ、オーストラリア、その他
 - NDL – 平成24年度書誌調整連絡会議(10/12開催)において方針表明(RDAと新書誌フレームワーク)
- ただし、たんなる交換用フォーマットの変換から、書誌データ作成方式およびシステムの全面的再構築までの、さまざまな可能性が考えられる

LC-BIBFRAMEの高次モデル(主クラス)

- 著作 (Work)
 - ❖ 目録対象資料の概念的本質を反映した資源
- インスタンス
 - ❖ その著作の、個別のものとしての具象化を反映する資源
- 典拠
 - ❖ 著作とインスタンスに現れた関係性を定義した主要な典拠概念を反映する資源。人名、地名、件名、組織名等
- 注釈
 - ❖ 追加の情報によって他のBIBFRAME資源を修飾する資源。図書館の所蔵情報、表紙の絵、レビュー等

Library of Congress. *Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Service*. 2012.11, 42 p.

<http://www.loc.gov/marc/transition/pdf/marclid-report-11-21-2012.pdf>

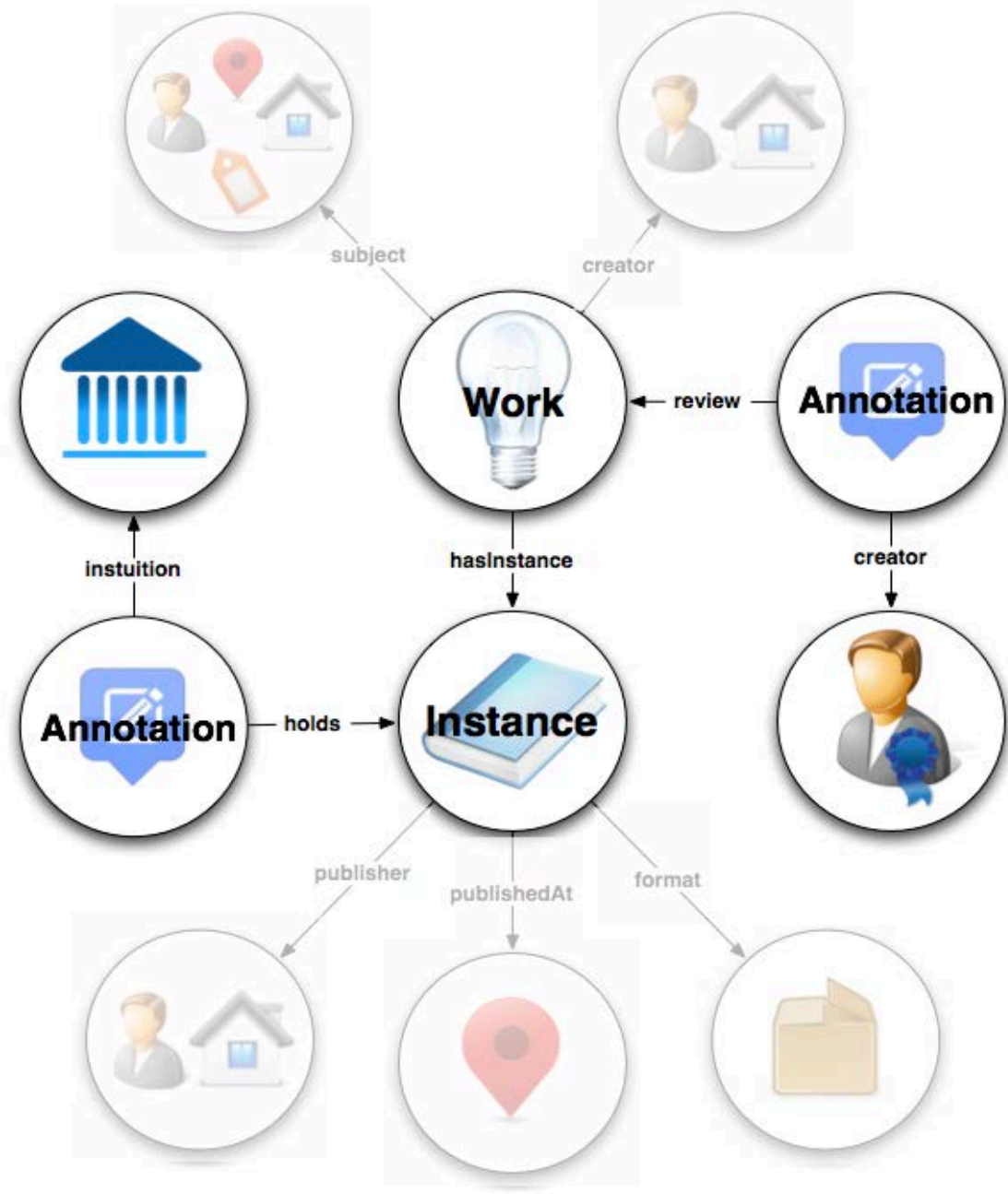


「著作資源とインスタンス
資源間の関係性を定義
するBIBFRAMEリンクト
データ・モデルとそのウェ
ブによるアドレス指定が
可能な典拠情報源への
文脈化の図式表現」

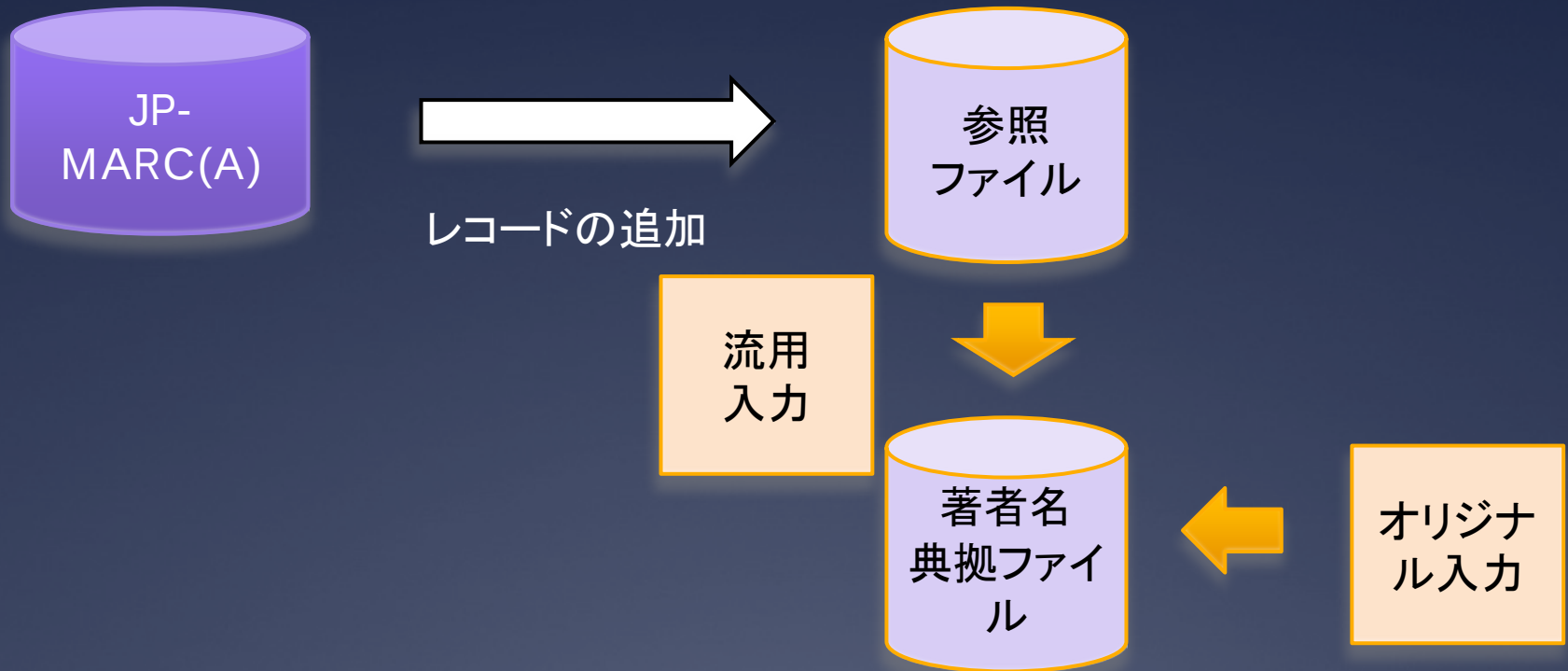
Library of Congress.
*Bibliographic Framework as a
Web of Data: Linked Data
Model and Supporting Service.*
2012.11, 42 p.
<http://www.loc.gov/marc/transition/pdf/marclid-report-11-21-2012.pdf>

「柔軟な注釈フレームワークの文脈でのBIBFRAME リンクトデータ・モデルの 図式表現」

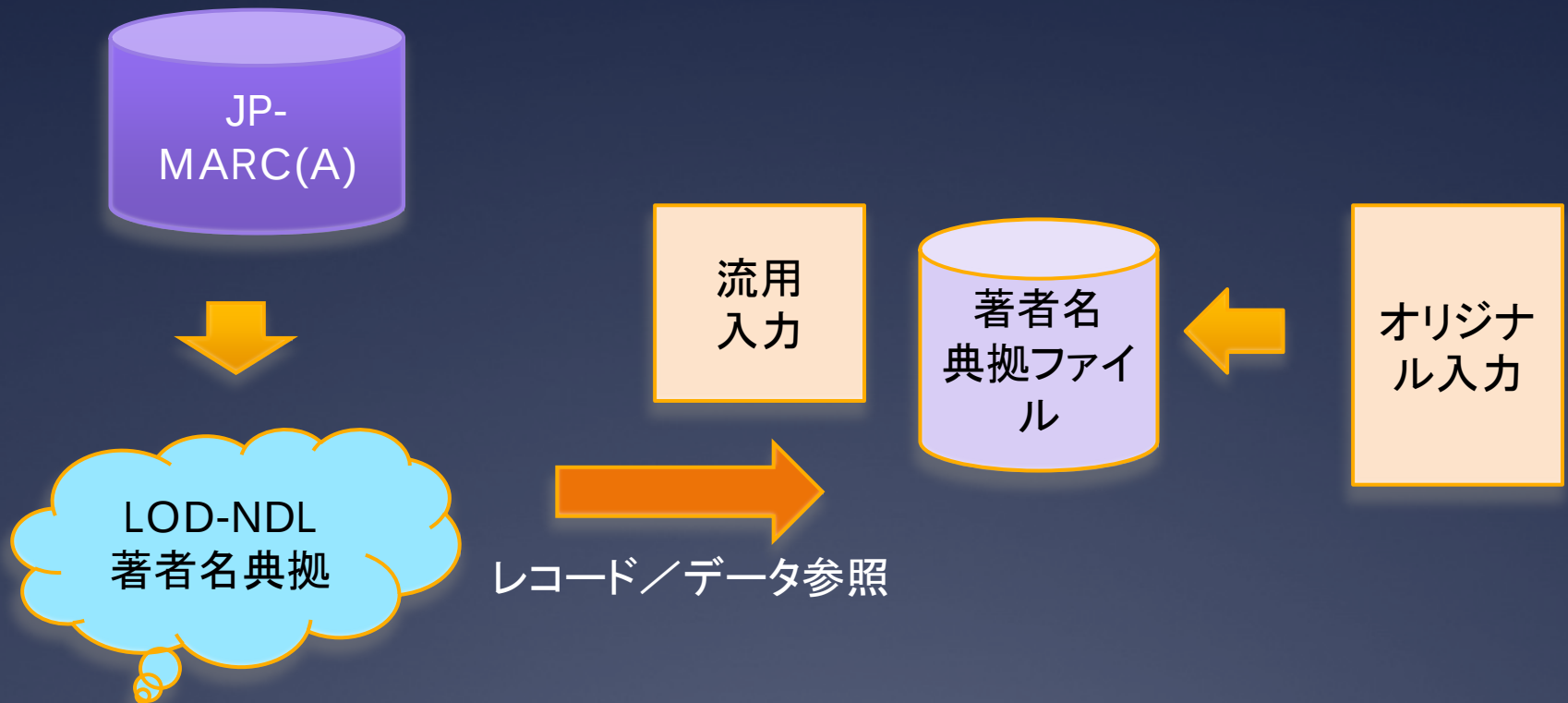
Library of Congress.
*Bibliographic Framework as a
Web of Data: Linked Data Model
and Supporting Service.*
2012.11, 42 p.
<http://www.loc.gov/marc/transition/pdf/marclid-report-11-21-2012.pdf>



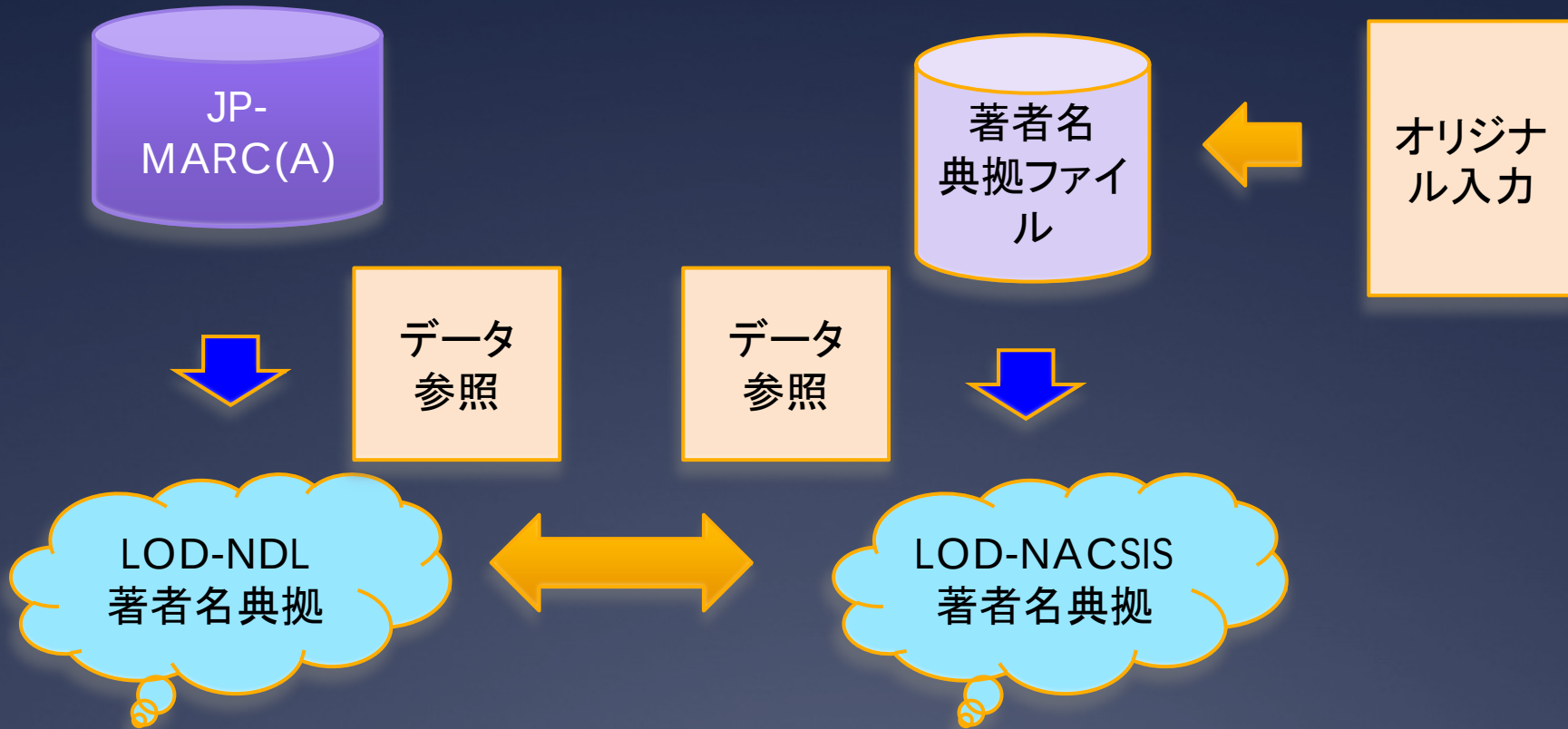
現行方式



仮想a



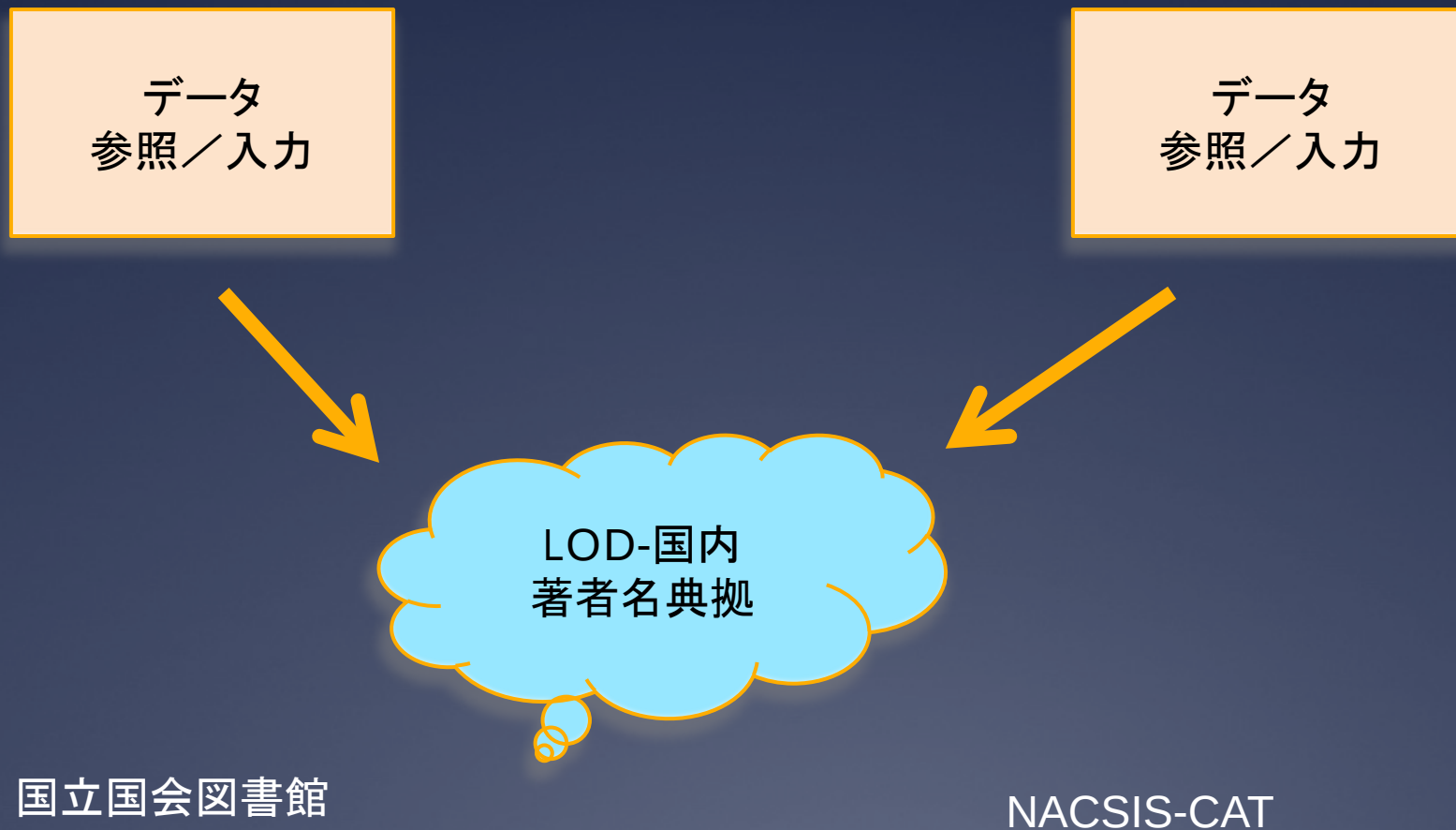
仮想b



国立国会図書館

NACISIS-CAT

仮想c



LOD対応の課題

❖ LODのメリットを享受するためには、対応するデータセットの公開が前提となるのは当然

1. データ形式の策定とこれに対応するシステム整備
2. データの同定識別のための識別子の設定および管理
3. 重複データ排除のための（海外を含む他機関との）連携
4. データの権利関係の定義（CC0 or ODC-BY？、参加機関との調整）
5. NIIと大学図書館の双方における、目録作成／検索機能提供の役割の再定義？

データの権利に対するアプローチ

- CC0 1.0 Universal – (Creative Commons) Public Domain Dedication
 - 著作権を主張しない
 - 多くの国立図書館、大学図書館が採用
- ODC-BY - (Open Data Commons) Attribution License released
 - 「(帰属)表示」のみ(「共有」「創作」「翻案」可能、「継承」「キープ・オープン」は不要)
 - OCLCが採用



4. これから何が必要か

- 国立情報学研究所 学術コンテンツ運営・連携本部 図書館連携作業部会報告書 『電子的学術情報資源を中心とする新たな基盤構築に向けた構想』
2012.3, 36 p.

http://www.nii.ac.jp/content/archive/pdf/content_report_h23_with_glossary.pdf

今後の基盤構築の方向性および原則

1. 電子情報資源の確保

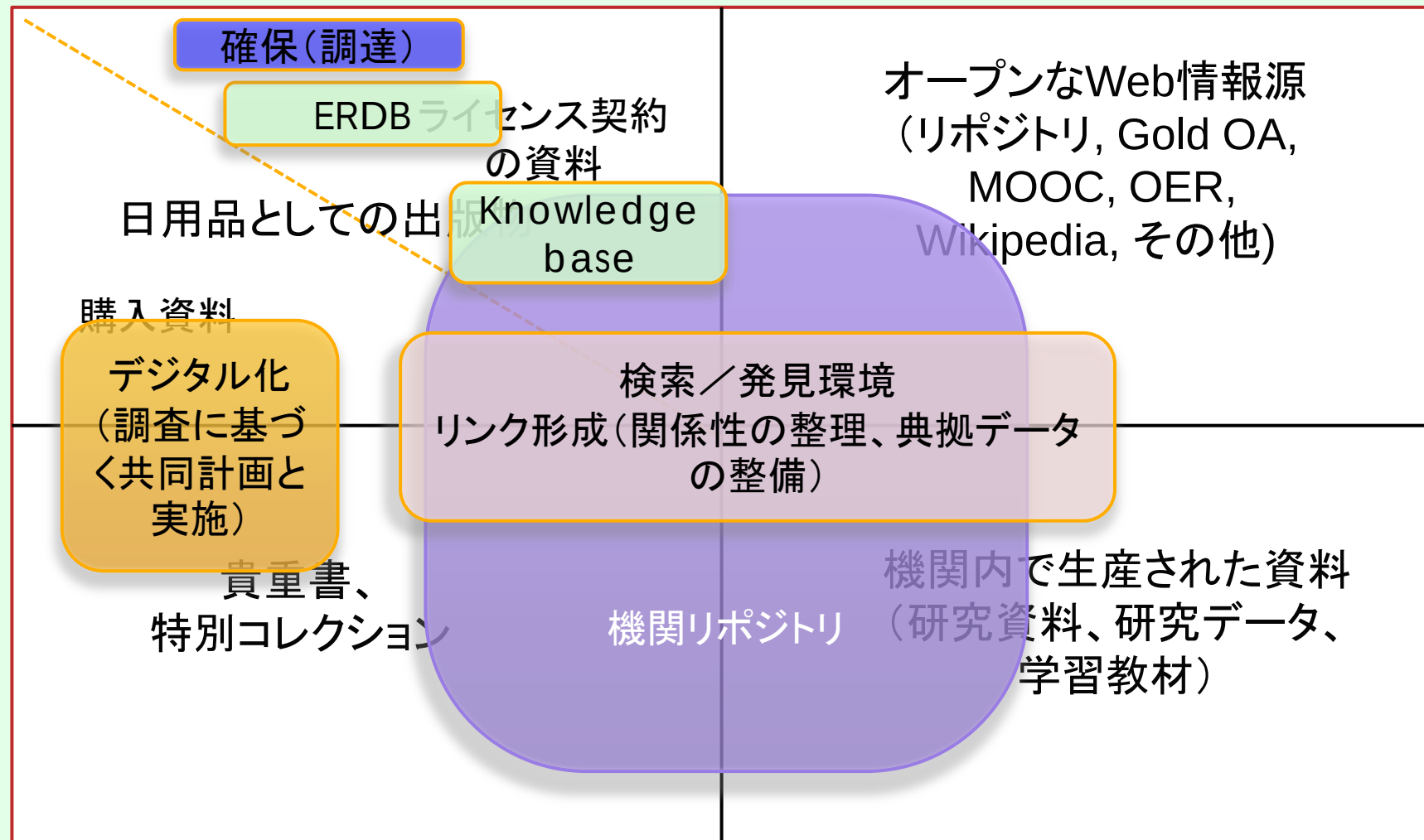
- 広範囲な情報源へのアクセスのための「ライセンス」の確保、既所蔵資料の電子化、効率的・効果的な提供を実現する仕組みの確立

2. メタデータのオープン化と相互接続性の確保

- 目録や論文のメタデータのオープン化への対応（APIの運用や識別子の整備に関する取り組みのより一層の推進）、組織面・制度面での整備（例. メタデータの知的所有権など）

3. 統合的発見環境とシステム基盤

- 利用者を良質な学術コンテンツに導くための網羅的な「発見」の仕組みの構築、電子情報資源と印刷体の両方について情報の最終ターゲットまで確実に到達できる環境（システム基盤を含む）の整備



統合インデックス

持続性の確保のための体制整備

1. 集中的な処理の仕組みの確立

- 電子情報資源のメタデータ形成
- LOD対応 - 典拠データ整備

2. 「国立情報学研究所と国公立大学図書館協力委員会との間における連携・協力の推進に関する協定書」(以下、協定書)の趣旨に基づく連携・協力事業としての基盤構築の推進

3. 国内外の諸機関間および事業との連携重複作業の排除およびグローバルな電子的情報基盤の構築への積極的な貢献

電子コレクション構築のための整備

- 既存資料の電子化

- 学術資料電子アーカイブの形成と利用の促進等

- ライセンシング（契約）

- 大学図書館コンソーシアム連合（JUSTICE）による合理的なライセンス取得、日本語学術図書（デジタル版）のライセンス契約の検討、ライセンス契約に伴う費用負担に関する検討および調査

- 目録データの整備

- 印刷資料の目録システムとしてのNACISIS-CATシステムの維持、LODへの対応、電子情報資源（ライセンス資料）のメタデータ、統合インデックスの構築・提供等

- オープンアクセス

- 学術情報のオープンアクセス化の推進（SPARC Japan、DRF、およびグローバルな枠組みとの連携を前提）

システムの整備

1. 電子リソース管理データベース(ERDB)

- ライセンシングにより利用可能となる電子リソースの書誌データ、契約データ、利用統計データ等を一元的に管理し、共同利用するためのシステムの構築
- “knowledge base”

2. 統合インデックス

- 国内外のさまざまな粒度の学術情報資源を網羅的、効率的に発見し、それらにアクセスするためのサービス基盤としての統合インデックスの整備