*RDA* 東京大会における議論を踏まえた研究データ共有の最新動向
研究データとオープンサイエンスフォーラム

# RDA東京大会参加報告

千葉大学附属図書館利用支援企画課　三　角　太　郎

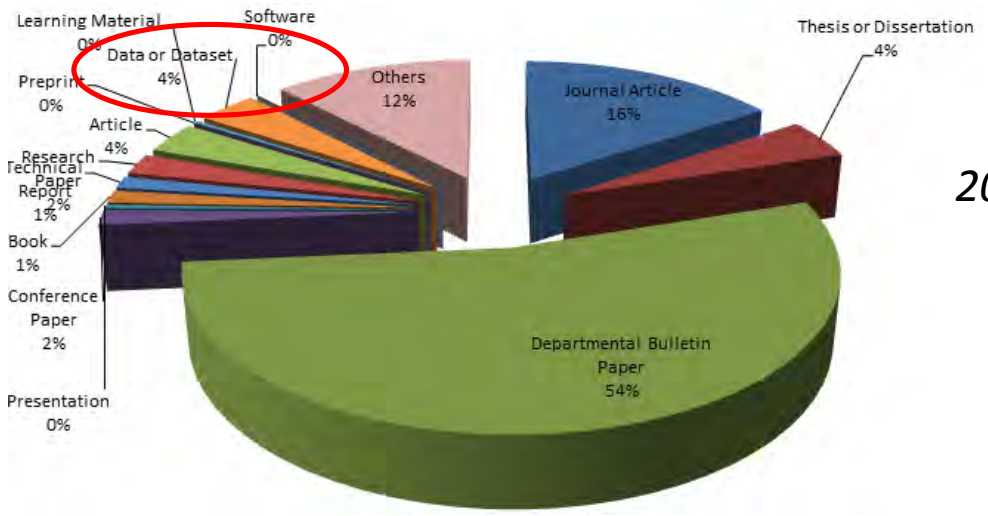
# 簡単に自己紹介

- 大学図書館員として十数年、閲覧から受入目録、雑誌、ILL、システム、古文書までたいていの業務は担当してきました。
- 三年ほど研究協力系の部署にいたこともあります。
- 昨年度から、オープンサイエンスにかかわっていますが、大学図書館として何ができるのか、図書館員に研究データが扱えるのか日々悩んでいます。
- 以下、少し長い前置き

# 自信がなくなる数字

99.6%

# Data set in IRDB(Institutional Repositories DataBase)



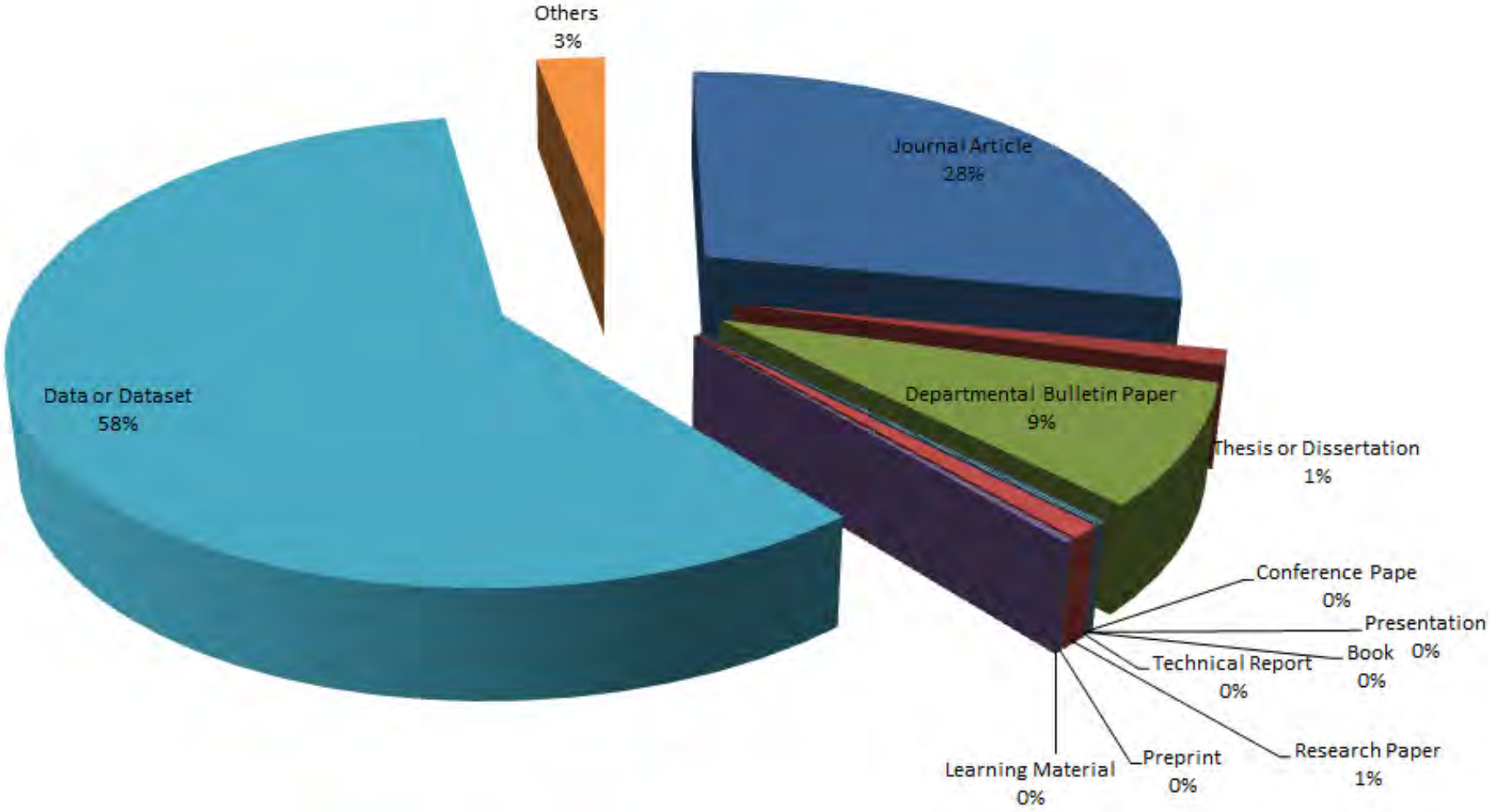

*2015.1*

**Total 52,421**

**Chiba Univ 52,223**

**52,223÷52,421 = 99.6%**

# Content t y pe (Chiba Univ.)




*2015.1*

# e-Science Project

**Surveys and research concerning the possibility of collaboration between e-Science projects and IRs**

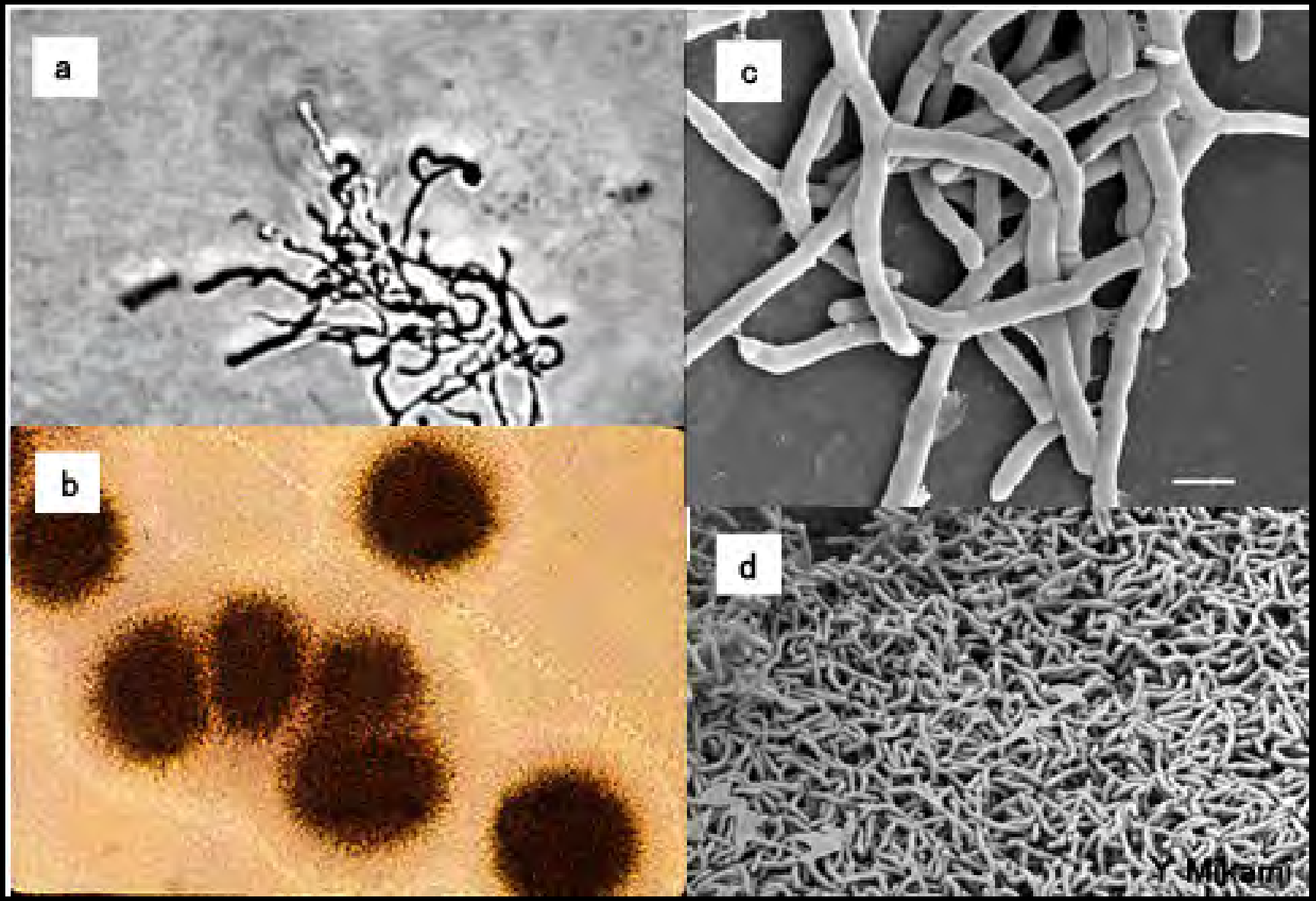
Lead institution: Chiba Univ.

Contributing institution: Kanazawa Univ. Kyushu Univ. Hokkaido Univ. Osaka Univ.

・Tagging images with interoperable metadata suitable for uses by various communities of researchers and students
・Development and appraisal of the platform system
・Effective and efficient management system for intellectual inputs and outcomes for research and higher education

# Contents Design

# Dataset



# Dataset

JH018725
HERBARIUM

JH022283

JH045192

JH000129

JH025050

JH031753

# 日本の機関リポジトリの実態

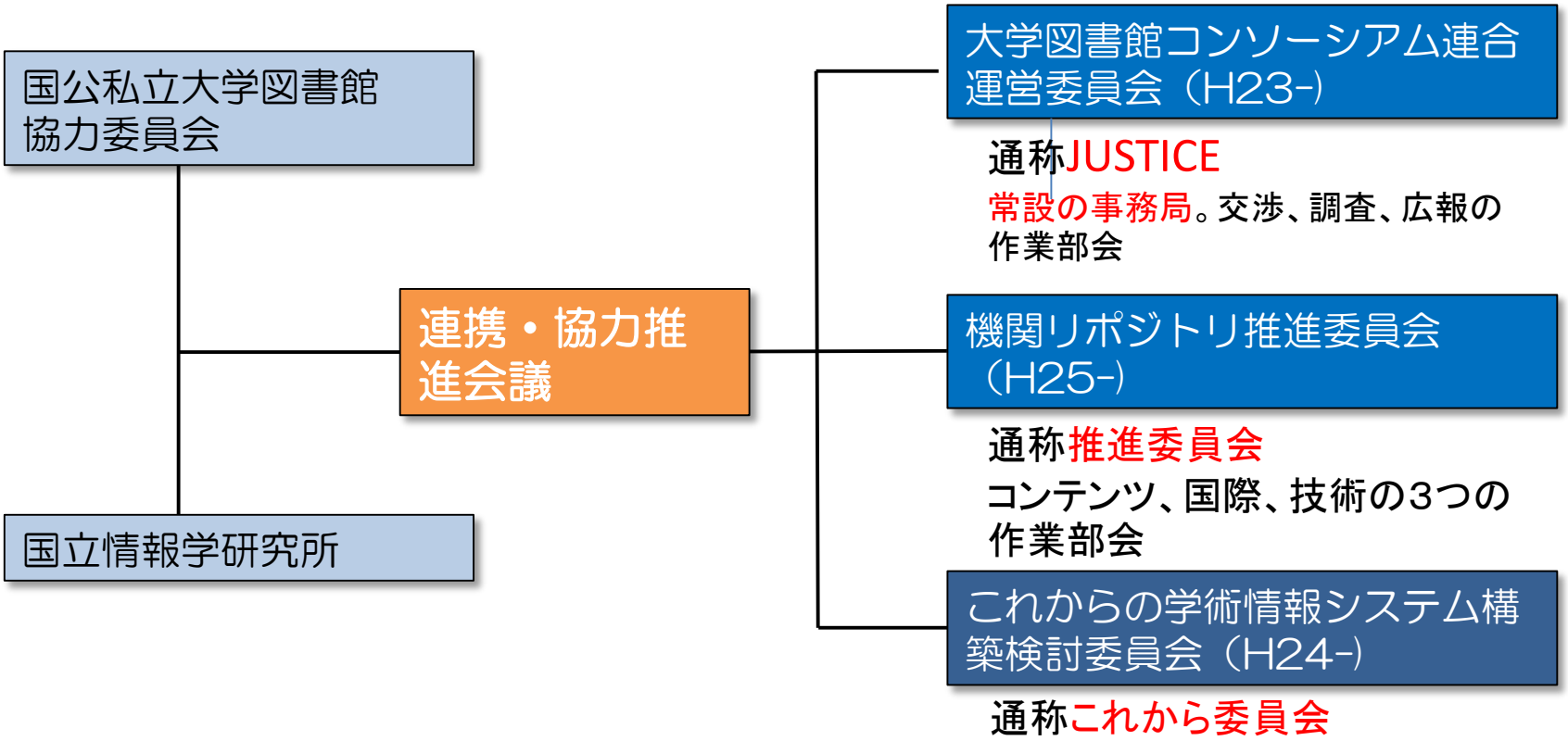
- 千葉大の登録している５万件のデータは、「研究データ」というよりも「標本データ」(もちろん標本データも重要である）。
- 日本国内の機関リポジトリでは、研究データを扱った実績はないに等しい。
- 「オープンサイエンスの潮流」がきているとは思うが、日本の大学図書館はそこに乗れるのか？

# 大学図書館全体のとりくみ

## 機関リポジトリ推進委員会

https://ir-suishin.repo.nii.ac.jp/

「大学共同利用機関法人情報・システム研究機構国立情報学研究所と国公私立大学図書館協力委員会により設立、「機関リポジトリを通じた大学の知の発信システムの構築」に関する事項を企画・立案し、学術情報の円滑な流通及び発信力の強化にかかる活動を推進することを目的とする。




http://www.nii.ac.jp/content/cpc/org/を基に作成

# オープンサイエンス班の活動計画

【1】OAポリシーの策定支援ツールの開発
【2】RDMトレーニングツールの開発
【3】研究データ対応メタデータスキーマの検討
【4】ケーススタディによる研究データ管理ノウハウの蓄積

# オープンサイエンスのために何をしなければならないか？

内閣府の報告書や海外の状況を自分なりに分析してみたが・・・

1. 公的助成の成果のオープンアクセス化
   → 論文等の研究成果情報の可視化

2. 研究データへの対応
   → オープンアクセスと研究倫理の二つの側面あり

# オープンサイエンスのために何をしなければならないか？

ヒト　図書館員のトレーニングコースの整備

カネ　大学の事業とするためのOAポリシー策定支援
　　助成団体の動向の調査
　　→　可能であれば、こちらから提案

モノ　システムの整備
　　1）　サーバー　今年度はとりくまない
　　2）　メタデータ
　　→　流通の基本はメタデータだが、現在のメタデータスキーマのJunii2では、助成情報が入らないし、著者情報も不十分

# ここまでが前置き

・日本では手探りで進めているところだが、世界ではどう動いているのだろう？

・世界の最前線が、日本に来てくれる、これはまたとない機会！

# 参加したセッション

- Breakout 1 -IG Education and Training on handling of research data: Focus on Research -Data handling related competences and skills

- Breakout 2 - Working Meeting Session 2　Joint meeting of IG Education and Training on handling of research data, WG RDA/CODATA Summer Schools in Data Science and Cloud Computing in the Developing World: International Coordination of Research Data Education and Training Requirements

- Breakout 3 - Working Meeting Session 3　IG Archives and Records Professionals for Research Data

- Breakout 4 -Joint meeting of IG Data Rescue, IG Data Fabric, IG Preservation e-Infrastructure, IG Domain Repositories, IG Libraries for Research Data: Rescuing, Re-Using and Sharing Data At Risk

- Breakout 5 -BoF on research data repository interoperability

# IG Education and Training on handling of research data: Focus on Research -Data handling related competences and skills

https://rd-alliance.org/node/971/all-wiki-index-by-group

We agreed at Plenary 5 (San Diego, March 2015) that we would draw together information on the skills and competences currently or soon likely to be required for handling research data in the following four professional areas:

それぞれの職種で必要なスキルを定義する

1. research librarians;
2. research administrators;
3. research infrastructure managers / operators;
4. researchers.

For each of these four professional areas, please contact the named person to contribute skills or competences that are considered essential or desireable in your own professional context. Please also contribute citations to any relevant resources or reading material which helps us to identify current skills and competences for research data handling.

# Essential skills 1

1.1 Essential skills:

Metadata (CE); Familiarity with various metadata standards;

Familiar with different data documentation practices, not just metadata schemas (AN)

Familiarity with various data standards (LW)

Data management planning (CE)

Data repositories (CE); knowledge of repositories and other digital initiatives (LW) Understanding the general requirments for preservable data or specific requirements for domain repositories (AN)

Data preparation (CE)

Marketing services (CE)

Building collaborations (CE); interaction and cooperation with business and IT personnel (LW)

Data backup and security (CE)

# Essential skills 2

Understand and create organisational culture change (LW)

Experience of data and document management (LW)

Training, advocacy and awareness-raising activities (LW)

Communicating the importance of data stewardship, ability to make the case (AN)

Managing and influencing business change activities (LW)

Experience of working with an HE research establishment (LW)

Understanding of RDM policy issues and best practice (LW) Funder requirements for data planning, handling, deposition (AN)

Familiarity with RDM lifecycle (LW)

Ethical and legal requirements in handling data from human participants (VVdE) and other sensitive data types (AN)

Review and appraisal of research data (VVdE)

Data sharing/distribution/publication (AN)

Intellectual property and Data licenses (AN)

Defining the working space: What is research data? (AN)

Migration between working file formats and preservation file formats (AN)

# Essential skills 3

File naming, organization, and versioning (AN)

Familiarity with existence of available tools for data handling (AN)

Identifying storage strategies for different data types, stages, quality. Selecting data for presearvation (AN)

Understanding of research reproducibility (AN)

From the Data Information Literacy (DIL) Program (AN)

- Discovery and Acquisition
- Ethics and Attribution
- Metadata and Data Description
- Cultures of Practice
- Data Management and Organization
- Data Curation and Reuse
- Data Quality and Documentation
- Data Processing and Analysis
- Data Visualization and Reuse
- Databases and Data Formats
- Data Conversion and Interoperability
- Data Preservation

# 気が遠くなりそうだが・・・

Research　Librarians : essential skill (so far)

まだまだ先は長い、とプレゼンターの方も・・・

# Wiki ページ

1.2 Desirable skills:

- Data citation (CE) (Would suggest this as essential - AN)
- Data literacy (CE)
- Data visualisation (CE)
- Data mining (CE)
- Databases (CE)
- ...

1.3 Background reading (currently in no particular order):

a) (LM) DigCurV competence framework for digital curation in cultural heritage organisations (including libraries and universities):

- Executive summary of the project and overall framework development process: http://www.digcur-education.org/eng/Resources/D2.1-Executive-Summary
- Version of the framework for practitioners: http://www.digcurv.gla.ac.uk/assets/downloads/dcsf_practitioner_lens.pdf
- Version of the framework for (middle) managers: http://www.digcurv.gla.ac.uk/assets/downloads/dcsf_manager_lens.pdf
- Version of the framework for senior executives: http://www.digcurv.gla.ac.uk/assets/downloads/dcsf_executive_lens.pdf

b) (LM) RDM Rose training resource v.3 (2015): http://rdmrose.group.shef.ac.uk/?page_id=1061

c) I (LM) did a keynote on this topic (RDM for research librarians including available training and guidance) to the association of French academic library directors in late 2013.

- Slide (no. 21) from this presentation distils the findings of the RILADS report http://www.slideshare.net/lm_hatii/nn-molloy-adbumrdoverview20130919slideshare(the full RILADS report is available here: https://rilads.wordpress.com/2013/06/10/rilads-report/if you want the detail)
- Slides 21 – 24 cover librarians skills for data management more widely: http://www.slideshare.net/lm_hatii/nn-molloy-adbumrdoverview20130919slideshare

c) (LM) Sheila Corrall's work is very useful in this area

情報源を丁寧に集めているのは、さすが！と思った。

# 所感

- まだ、 Essential skillsを集めている段階。
- コンタクトパーソンを明示しているのは良いと思った。
- 情報の量、そして、それを丁寧に集めているのはさすが！と思った。
- 正直言って消化不良であり、おそらく、これをそのまま日本に適用するのも困難だが、図書館員に求められているスキルがどういうものかはつかめてくるであろう。

→　ちゃんと分析しなければ！

# BoF on research data repository interoperability

- The Research Data Repository Interoperability Working Group will establish standards for interoperability between different research data repository platforms. These standards may include (but are not limited to) a generic API and import/export formats

*Removing Barriers*

*Easier Collaboration*

*Creating Commonalities*

# BoF on research data repository interoperability

Engagement with Existing Work

A number of related standardization efforts have already taken place; for example, the OAI protocol for metadata harvesting, the SWORD protocol for repository deposits, and the re3data.org schema for collecting information on research data repositories for registration. The Research Data Repository Interoperability WG will review these and other related standards to see how they might be adopted or extended to support our goals. This review period will ensure that we do not duplicate existing efforts.

- OAI-PMH
- SWORD Protocol
- Re3data Schema
- Portland Common Data Model
- METS
- Linked Data Platform (LDP)
- Dariah Storage API
- Resource Sync

https://docs.google.com/document/d/11XZBXIxSOE_d0n1JsaYx2LhRwkwel5OQNHBaXKG-a_o/edit

# BoF on research data repository interoperability

Representatives of several major repository platforms have already joined this working group, including:

- DSpace
- Fedora
- Hydra
- Metacat
- KIT Data Manager
- Dash research data portal/Merritt
- Dataverse

JAIRO Cloud
も手をあげなければ！

These representatives have agreed to implement the standards recommended by the Research Data Repository Interoperability WG in their respective repository platforms. We will continue to seek representatives from a variety of repository platforms and services to ensure that this working group's deliverables are widely adopted.

# IG Libraries for Research Data

- **Libraries for Research Data (L4RD) Interest Group Meeting**

https://rd-alliance.org/ig-libraries-research-data.html

RDA
RESEARCH DATA ALLIANCE
UNITED STATES

## 23 Things: Libraries for Research Data

*An overview of practical, free, online resources and tools that you can begin using today to incorporate research data management into your practice of librarianship.*

*Research Data Sharing Without Barriers*

### Learning Resources

Librarians are learning how to apply the principles of library science to solve problems and to provide new services related to research data.

1. A **"top ten" list of recommendations** for libraries to get started with research data management from LIBER, http://bit.ly/1qUvKG3
2. Relevant concepts are presented and mapped in the **e-Science Thesaurus**, http://bit.ly/1LEo4h8
3. Understanding the life of research data with the **DCC Curation Lifecycle Model**, http://bit.ly/1MoGGGv
4. **MANTRA** online training modules for librarians,

Learning Resources
Data Reference and Outreach
Data Management Plans
Data Literacy
Citing Data
Data Licensing and Privacy
Digital Preservation
Data Repositories
and a Community of Practice

***..to help librarians engage in research data management!***

### Data Reference & Outreach

Librarians are answering questions about data from patrons and conducting outreach to assess the data needs of their researchers and students.

7. Begin a conversation with a researcher about data by

10. Questions about data answered by experts on the **DataQ** forum, http://bit.ly/1MoH4Vq

### Data Management Plans

Librarians are becoming familiar with funder requirements and consulting with researchers to help them write and implement effective data management plans.

11. One example is the **DMPTool** that lists funder requirements in the United States and builds a plan by asking the researcher to answer a series of questions. Other countries such as the U.K. and Canada have similar tools, http://bit.ly/1LuNZMH

### Data Literacy

Librarians are including data in their information literacy instruction, to

# 「研究データの図書館」からの２３のアドバイス

RESEARCH DATA ALLIANCE
UNITED STATES

## 「研究データの図書館」からの 23 のアドバイス

ライブラリアンシップの実践で研究データ管理をするために、即座に使える実践的無料オンラインリソースとツールの概要

**障壁なきデータの共有 *(Research Data Sharing Without Barriers)***

**学習リソース**
図書館員は研究データに関連したソリューション、そして新たなサービス提供のために、図書館学の原理をいかに応用するかを学んでいます

1. LIBER から研究データ管理を始めようとしている図書館へのアドバイス「トップ・テン」 http://bit.ly/1qUvKG3
2. 関連するコンセプトが e-Science Thesaurus に紹介・解説されています http://bit.ly/1LEo4h8
3. 研究データのライフサイクルについては DCC Curation Lifecycle Model をご覧ください http://bit.ly/1MoGGGv
4. 図書館員のための ANTRA オンライン・トレーニング・モジュールをご活用ください http://bit.ly/1RRvVju

データレファレンスとアウトリーチ
データ管理計画
データリテラシー
メタデータ
データ引用
データライセンスとプライバシー
デジタル保存
データレポジトリ
実践コミュニティ

**データレファレンスとアウトリーチ**
図書館員は、利用者から寄せられたデータに関する質問に対する回答や、研究者や学生のデータニーズについて調べるためのアウトリーチを行っています

7. Conducting a Data Interview を用いて、研究者とデータに関する対話を始めましょう http://bit.ly/1opHoyQ

て理解を深め、効果的なデータ管理計画を作成し、実施するために、研究者に意見を聞いています

11. 米国内の資金提供機関の必要条件がリストしてあり、研究者に一連の質問に答えてもらうことで計画を立てる、DMPTool が例として挙げられます。イギリスやカナダなどの他国にも同様のツールが存在します http://bit.ly/1LuNZMH

**データリテラシー**
図書館員は情報リテラシー教育に「データ」を組み入れ、利用者にとっていつデータが必要になるのか、また利用者がデータを見つけ、評価し、使用する能力がいつ必要になるのかを知る手助けをしています

12. データ・情報リテラシー・プロジェクト（The Data Information Literacy

研究データ管理への第一歩

https://rd-alliance.org/system/files/documents/23Things_Libraries_For_Data_RDA%EF%BC%BFjp.pdf

# Purdue大学の例

*Michael Witt. March 3, 2016. Institutional Data Repository Services at Purdue. Digital Repository Federation, National Institute of Informatics, Tokyo, Japan.* → *RDA*関連イベントとして開催。

図書館員の研究データ管理へかかわるタイミング

1) data management plans に対する支援
2) a new project is created or a grant is awarded
3) consult or collaborate on project if needed
4) submitted datasetsに対するレビューと投稿
5) 10年の保管期間の後に、図書館のコレクションに加えるかどうかを選定

主題分野に対する知識もある程度必要か？

# 所　　感

- どのセッションでも、ImpactやNextStepが明示される→見習うべき
- 参考情報源は、すでにたくさんある
- Purdue大学の例を聞いたが図書館員の層が厚く、専門性も高い

  →理系のバックグラウンドをもった図書館員も多数

**日本の図書館は対応できるのか？**

# しかし図書館がやらねばどこがやる？

○　学術情報のアーカイブをミッション

→　大学内でアーカイブの文化をもつ
組織は図書館・博物館・記念室ぐらい。

○　書誌情報（メタデータ）作成の
ノウハウをもっている

→　図書館員ならある程度の
書誌作成のスキルはもっている

# 今後の方針

- スモールスタート
- 図書館ができる事とできない事を明確に！
- 自力でのインフラ立ち上げが困難な分野がサービスの対象
- サイエンスという言葉に引きずられない
  →人文・社会科学も立派なサイエンス
- カレントの研究データよりも、まずは枯れたデータの長期保存。
- 国際的なスタンダードを常に意識

*Thank You*