

## シンポジウム・ワークショップ「大学における研究 データサービス」

Imker, Heidi  
University of Illinois at Urbana-Champaign

Mischo, William H.  
University of Illinois at Urbana-Champaign

Schlembach, Mary C.  
University of Illinois at Urbana-Champaign

Miyazaki, Yuta  
Kyushu University Library

他

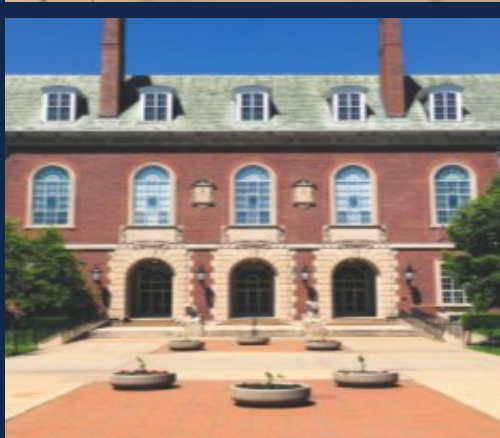
<https://hdl.handle.net/2324/2547228>

---

出版情報 : 2019-12-05. Department of Library Science, Graduate School of Integrated Frontier Sciences, Kyushu University

バージョン :

権利関係 :



# Data Management Support for Researchers in Practice

William H. Mischo

Mary C. Schlembach

Grainger Engineering Library Information Center  
University of Illinois at Urbana-Champaign, U.S.A.

December 5, 2019

**I** ILLINOIS  
University Library

# 研究者に対する データマネジメントサポート の実践

William H. Mischo

Mary C. Schlembach

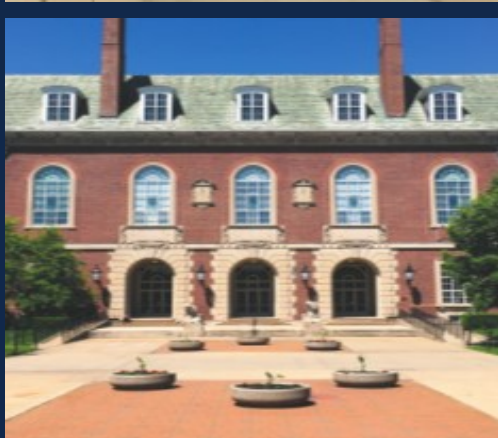
Grainger Engineering Library Information Center  
University of Illinois at Urbana-Champaign, U.S.A.

December 5, 2019

**I** ILLINOIS  
University Library

日本語訳：九州大学附属図書館eリソース課 宮崎祐汰

\*このスライドの日本語訳は、シンポジウムの配布資料として作成したものです。  
専門家による訳ではないため、誤りがある場合があります。



## Researcher Views on Data Sharing

- Studies show that U.S. and Japanese researchers have an awareness of data sharing practices and an acceptance of the value of sharing data.
- Our studies and interviews have shown that researchers treat data as one of many necessary elements in the larger scholarly communication workflow.
- Researchers view data management as part of the overarching research lifecycle, not as a separate process.

# データ共有における研究者の視点

- 研究により、米国と日本の研究者はデータ共有の実践を意識し、データ共有の価値を認めていることが分かっている。
- 我々の研究とインタビューにより、研究者はデータをより大きな学術コミュニケーションのワークフローにおける、多くの必須要素の一つとして扱うことが分かっている。
- 研究者は、データマネジメントを独立したプロセスとしてではなく、包括的な研究ライフサイクルの一部として捉えている。

## The Conduct of Research

- The knowledge creation and scholarly communication process have changed.
- The growth of electronic journals, books, and conferences has changed the medium of knowledge dissemination.
- The rise of grid and cloud computing; big data; simulation and modeling software; artificial intelligence, machine learning and deep learning; and data visualization and analytics tools have changed the face of research.

# 研究の遂行

- 知の創造と学術コミュニケーションプロセスは変化した。
- 電子ジャーナルや電子ブック、電子会議録の進展により、知を広める媒体が変化した。
- グリッドコンピューティング及びクラウドコンピューティング；ビッグデータ；シミュレーション及びモデリングソフトウェア；人工知能、機械学習及びディープラーニング；データの可視化ツール及び分析ツールの台頭により、研究の様相が変化した。

# Data Management Mandates and Services

- Data Management Plans (DMPs) became a required element in U.S. National Science Foundation (NSF) proposals in January 2011.
- DMPs have been required for some U.S. National Institute of Health (NIH) grant proposals and are likely to be required for all proposals soon.
- Researchers in the U.S. have not yet significantly changed their data management practices to conform with these federal government data sharing mandates.



# データマネジメントの義務及びサービス

- データマネジメントプラン(DMP) は、2011年1月に米国国立科学財団(NSF)の申請書において必須要素になった。
- DMPは、米国国立衛生研究所(NIH)のいくつかの助成金の申請書で必須とされており、すぐにもすべての申請書で必須とされる可能性がある。
- 米国の研究者は、これら連邦政府のデータ共有義務に準拠するためにデータマネジメントの実践の抜本的な変更をまだ行っていない。

# Data Management Services

- Establishing an effective data management service program presents challenges.
- There are known subject discipline differences in data management needs. This needs to be considered.
- Researchers in the U.S. often rely on graduate students for day-to-day management of data recordkeeping.
- Libraries have an opportunity to develop data management instructional programs and services that support the data workflow needs and practices of researchers.

# データマネジメントサービス

- 効果的なデータマネジメントサービスプログラムを確立するには課題がある。
- データマネジメントのニーズにおいては、主題分野別に相違があることが知られており、考慮する必要がある。
- 米国の研究者は、日々のデータ記録の管理を大学院生に任せていることがよくある。
- 図書館には、データマネジメントの教育プログラムとサービスを開発する機会がある。それらは、研究者のデータワークフローのニーズと実践を支援する。

# Instructional Program

- The changing knowledge creation environment gives libraries an opportunity to become closely involved in the research workflow.
- The focus at the University of Illinois Library has been on developing data management and curation instructional programs that can be integrated into presentations on other aspects of scholarly communication.
- Our data management presentations are directed at research groups and take place in classrooms and research labs.

# 教育プログラム

- 知を創造する環境が変化していくことで、図書館は研究ワークフローに密接に関わる機会を得た。
- イリノイ大学図書館で重視したのは、学術コミュニケーションの他の側面に関するプレゼンテーションに統合できる、データマネジメント及びデータキュレーションの教育プログラムの開発である。
- 我々のデータマネジメントプレゼンテーションは研究グループ向けで、教室や研究室で行われる。

## Pre-Data Management Instruction Checklist

- Before the data management instruction session, the librarian fills out a pre-data management instruction checklist.
- The checklist helps familiarize the instructor with the research focus of the particular group and better address the broad scholarly communication needs of the research group.
- Our data management instruction includes information on open access and open data, peer review, the selection of the best journals to publish in, and the available disciplinary repositories for dataset deposit.

# データマネジメント教育の事前チェックリスト

- データマネジメント教育セッションの前に、図書館員は、データマネジメント教育事前チェックリストを記入する。
- チェックリストのおかげで、インストラクターは、特定のグループの研究の焦点になり、かつ、研究グループの広範な学術コミュニケーション上のニーズにより注意を向けることができる。
- 我々のデータマネジメント教育は、オープンアクセスとオープンデータ、ピアレビュー、投稿するのに最適なジャーナルの選択、そして、データセットデポジットのために利用可能な分野別リポジトリについての情報を含んでいる。

# Pre-Data Management Instruction Worksheet

- Information about the faculty member(s):
  - Department and rank
  - Areas of research
  - Number of graduate students and post-docs
  - Major centers or consortia
- Publications by the faculty members in the group
  - H-Indexes and publications in Scopus
  - Data policies for publishers
  - Disciplinary dataset repositories (ISI Data Citation Index)
- Grants received by faculty member(s)
  - Funding agencies
  - Data policies of funders



# データマネジメント教育事前ワークシート

- 教員メンバーについての情報:
  - 部局と地位
  - 研究領域
  - 大学院生とポスドクの人数
  - 主要なセンター又はコンソーシアム
- 当該グループの教員メンバーによる出版物
  - Scopus中のHインデックスと出版物
  - 出版社のためのデータポリシー
  - 分野別データセットリポジトリ (ISI Data Citation Index)
- 教員メンバーが受け取った助成金
  - 助成機関
  - 助成機関のデータポリシー

# Six Stages of Data Processing

- Data management typically requires the manipulation of raw data to create processed data.
- We teach the six stages of data processing:
  - Collection processes
  - Preparation and manipulation
  - Input and conversion
  - Processing and manipulation
  - Output and interpretation
  - Storage and curation

# データ処理の6段階

- データマネジメントは、一般的に、処理済みのデータを作成するため生データを操作することが必要である。
- 我々は、以下の6段階のデータ処理を教えている：
  - 収集過程
  - 準備と操作
  - 入力と変換
  - 処理と操作
  - 出力と解釈
  - 保管とキュレーション

# Areas of Focus in Instruction

- Organization of data. Folder and file naming conventions.
- Documentation – file naming conventions, metadata header files, readme files, data dictionaries.
- Storage and backup, including the 3-2-1 rule for 3 copies of data, 2 kinds of media, and 1 remote location.
- Archiving time periods and embargo period.
- Sharing solutions and mechanisms.

# 教育における重点領域

- データの組織化。フォルダとファイルの命名法。
- 文書化 – ファイルの命名法、メタデータヘッダーファイル、Readmeファイル、データディクショナリー。
- 保管とバックアップ。3つのデータコピー、2種類の媒体、1つのリモートロケーションという、3-2-1ルールを含む。
- アーカイブ期間とエンバーク期間。
- 解決策とメカニズムの共有。

## Planning and Production Tools emphasized in Instruction Sessions

- The DMPTool and a custom data management plan template created by the Grainger Engineering Library at Illinois and used in over 300 NSF grant applications.
- We show a chart indicating which U.S. Federal agencies require a data management plan and which agencies require the deposit of articles produced under grant support.
- Description of Production Tools: Illinois Data Bank; IDEALS Institutional Repository; subject disciplinary repositories.

# 教育セッションで強調する計画ツール及び成果提供ツール

- DMPToolと、イリノイ大学のGrainger Engineering Libraryによって作成され、300以上のNSF助成金申請書において用いられている、カスタムデータマネジメントプランテンプレート。
- 米国のどの連邦当局がデータマネジメントプランを要求しているのか、及び、どの当局が助成金補助の下で生産された論文のデポジットを要求しているのかを示す表を公開している。
- 成果提供ツールの種類：Illinois Data Bank; IDEALS Institutional Repository; 主題分野別リポジトリ。

# Assessment and Grant DMP Analysis

- How successful has the Library DMP instruction program been and what effect has the Library's Research Data Service had on data management plans submitted by Illinois researchers?
- We have a lot of anecdotal evidence from our faculty that the data management services have been useful.
- In 2013 and also in 2017 we analyzed the data management plans submitted by University of Illinois researchers in NSF grants.



# 評価と助成金のDMPの分析

- 図書館のDMP教育プログラムは、どの程度成功しているのか。Library's Research Data Service（図書館の研究データサービス）は、イリノイ大学の研究者によって提出されたデータマネジメントプランにおいて、いかなる効果を持っているか。
- 教員から、データマネジメントサービスは有益であるという、個人的な見解に基づく根拠を多く得ている。
- 2013年及び2017年に、イリノイ大学の研究者によって提出された、NSF助成金のデータマネジメントプランを分析した。

# NSF Data Management Plans

- Working with the campus Office of Sponsored Programs and Research Administration (OSPRA), we were able to examine all grant proposals sent.
- We analyzed the data management plans in all submitted NSF grant requests, attempting to characterize and classify DMPs into categories
- We examined 1,260 DMPs from 2011 to 2013 (298 funded) and 869 DMPs from 2014 to 2015 (250 funded).

# NSFデータマネジメントプラン

- 学内のOffice of Sponsored Programs and Research Administration (OSPRA) と協働することで、送付された全ての助成金申請書を調査することができた。
- DMPを特徴づけ、カテゴリーに分類するために、提出された全てのNSF助成金申請のデータマネジメントプランを分析した。
- 2011年から2013年までの1260件のDMP（内298件採択）と2014年から2015年までの869件のDMP（内250件採択）を調査した。

# Reasons for DMP Analysis

- To try to determine if the Library's instructional and informational services were successfully influencing the DMPs that researchers were submitting.
- To determine what data storage and data sharing mechanisms and formats were being proposed in NSF DMPs.
- To determine what dataset repositories – including local campus repositories – were being proposed in DMPs.
- To determine if local DMP templates were being used in NSF grant proposals.

# DMPの分析を行う理由

- 図書館の教育及び情報サービスが、研究者が提出していたDMPに首尾よく影響を与えていたか否かを明らかにしようとするため。
- NSFのDMPにおいて、どのデータ保管場所と、データ共有のメカニズム及びフォーマットが提案されていたのかを明らかにするため。
- DMPにおいて、どのデータセットリポジトリ – ローカルキャンパスリポジトリを含む – が提案されていたのかを明らかにするため。
- ローカルDMPテンプレートがNSF助成金申請書で使われていたか否かを明らかにするため。

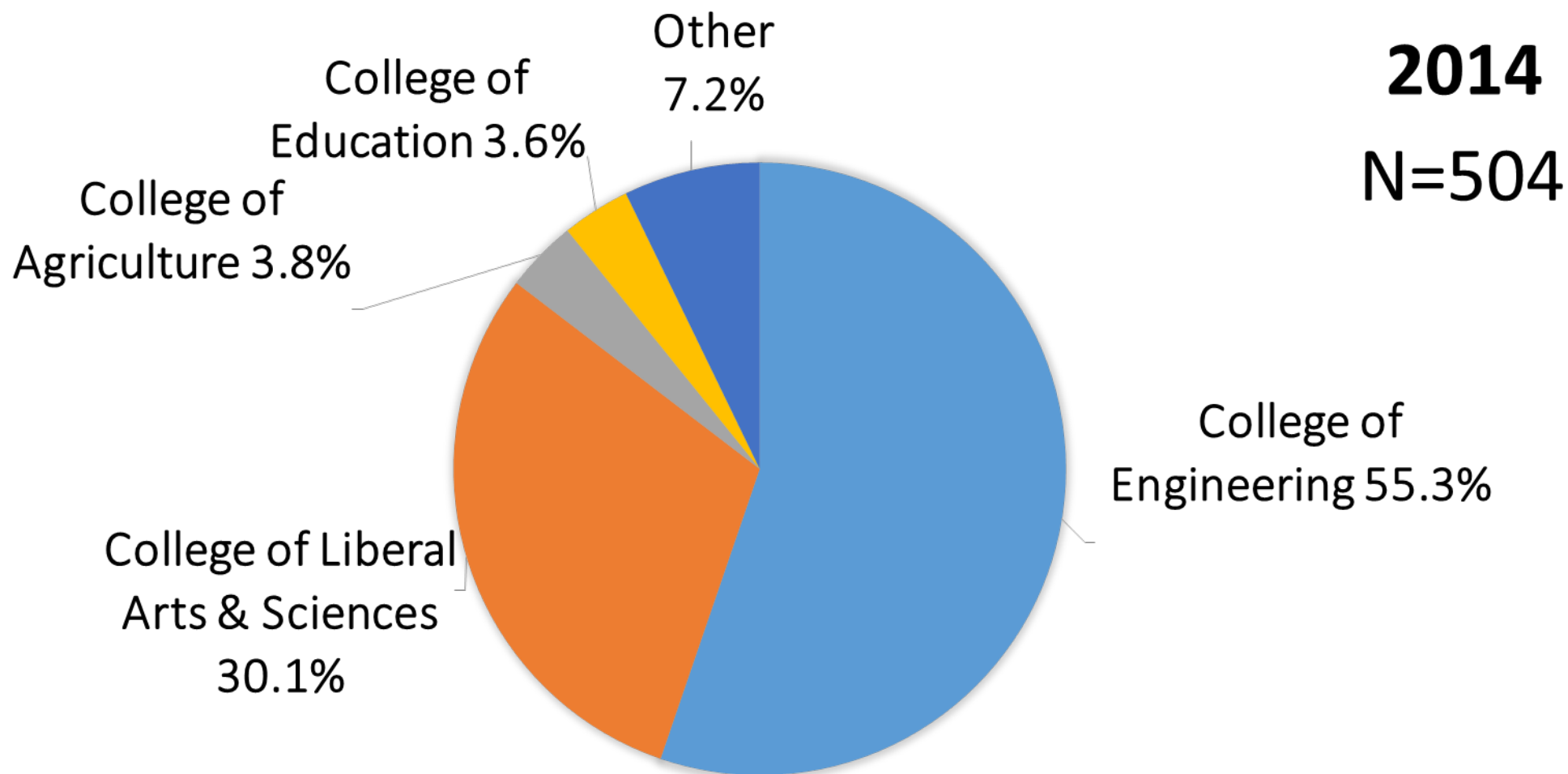
# DMP Categories

- Our analysis recorded the project dataset storage location(s):
  - Principal Investigator (PI) Server
  - PI website
  - Departmental server
  - Campus storage (Box) or campus cluster storage
  - Remote or Disciplinary Server or Repository
- We also recorded the dataset medium(s): PDF, ASCII, CSV, MatLab, Mathematica, GIS data, jpg, analog specimens, SQL datasets, software, etc.

# DMPカテゴリー

- 分析の結果、記録されたプロジェクトのデータセットの保管場所:
  - 研究代表者(PI)のサーバー
  - PIのウェブサイト
  - 部局のサーバー
  - キャンパスストレージ(Box)又はキャンパスクラスターストレージ
  - リモート又は分野別のサーバーやリポジトリ
- 記録されたデータセットの媒体: PDF, ASCII, CSV, MatLab, Mathematica, GIS data, jpg, analog specimens, SQL datasets, software, etc.

# DMP Categories





# DMP Storage Locations

2011 n=475		
Category	Number	Percent
PI Server	110	23.1%
PI Website	188	39.5%
Campus	155	32.6%
Department	82	17.2%
Remote	226	47.5%
Disciplinary	125	30.2%
Cloud	22	4.6%
Publication	336	70.7%
Analog	84	17.6%
Specimens	101	21.2%
Optical Disc	22	4.6%
Template Used	64	13.4%
Not Specified	139	29.2%
No Data	41	8.6%
Funded	127	26.7%

2014 n=504		
Category	Number	Percent
PI Server	178	35.3%
PI Website	223	44.2%
Campus	241	47.8%
Department	98	19.4%
Remote	239	47.4%
Disciplinary	162	32.1%
Cloud	50	9.9%
Publication	387	76.7%
Analog	118	23.4%
Specimens	99	19.6%
Optical Disc	18	3.5%
Template Used	120	23.8%
Not Specified	145	28.7%
No Data	26	5.1%
Funded	157	31.1%

July-Dec 2015 n=365		
Category	Number	Percent
PI Server	140	38.3%
PI Website	174	47.6%
Campus	164	44.9%
Department	97	26.5%
Remote	164	44.9%
Disciplinary	112	30.6%
Cloud	42	11.5%
Publication	299	81.9%
Analog	92	25.2%
Specimens	74	20.2%
Optical Disc	13	3.5%
Template Used	62	16.9%
Not Specified	105	28.7%
No Data	44	12.0%
Funded	93	25.4%

# DMP Analysis Conclusions

- We did not find any statistically significant differences (using chi-square) in proposed data storage locations between the unfunded and funded proposals.
- However, a higher percentage of the recently funded proposals propose a disciplinary repository as the data storage choice.
- We recommend that researchers consider the use of a disciplinary repository or our Illinois Data Bank repository for their datasets.
- The Library can play an important role in assisting researchers with data management plans and data storage/reuse.

# DMPの分析の結論

- 採択された申請とされなかった申請との間には、提案されたデータ保管場所において、統計的に有意な相違(カイ二乗を用いる)は見出せなかった。
- しかし、最近採択された申請は、高い割合で、データ保管の選択肢として、分野別リポジトリを提案している。
- 我々は、研究者が、データセットのために、分野別リポジトリ又はIllinois Data Bankリポジトリの利用を検討することを薦めている。
- 図書館は、データマネジメントプランとデータ保管／再利用に関する研究者への支援において、重要な役割を果たすことができる。