



OKAYAMA  
UNIVERSITY

# 米・コーネル大学出張報告

研究データ管理 etc.

2025/5/13 COAR Annual Conference 2025

岡山大学図書館 情報管理課

情報整備グループ 石丸 綾華

\*2025/5/26 Web掲載のためスライド一部修正

期間：2024年12月14日～20日

訪問先：コーネル大学（アメリカ）

- 1865年設立の私立大学  
メインキャンパス：ニューヨーク州イサカ
- アイビーリーグ所属  
THE世界大学ランキング20位
- 学部数：8  
大学院数：7
- 学生数：26,793人  
教職員数：11,674人



12月16日～17日：Olin Library, Mann Library（イサカ）

12月18日：Weill Cornell Medicine（マンハッタン）

メンバー：図書館長、電子ジャーナル契約担当 主査、  
機関リポジトリ担当、医学系分館職員、URA（計5名）



Cornell Data Services

Data Management Plan の作成からストレージ選定まで至れり尽くせり

研究を始めようかな（と研究者が思ったら）



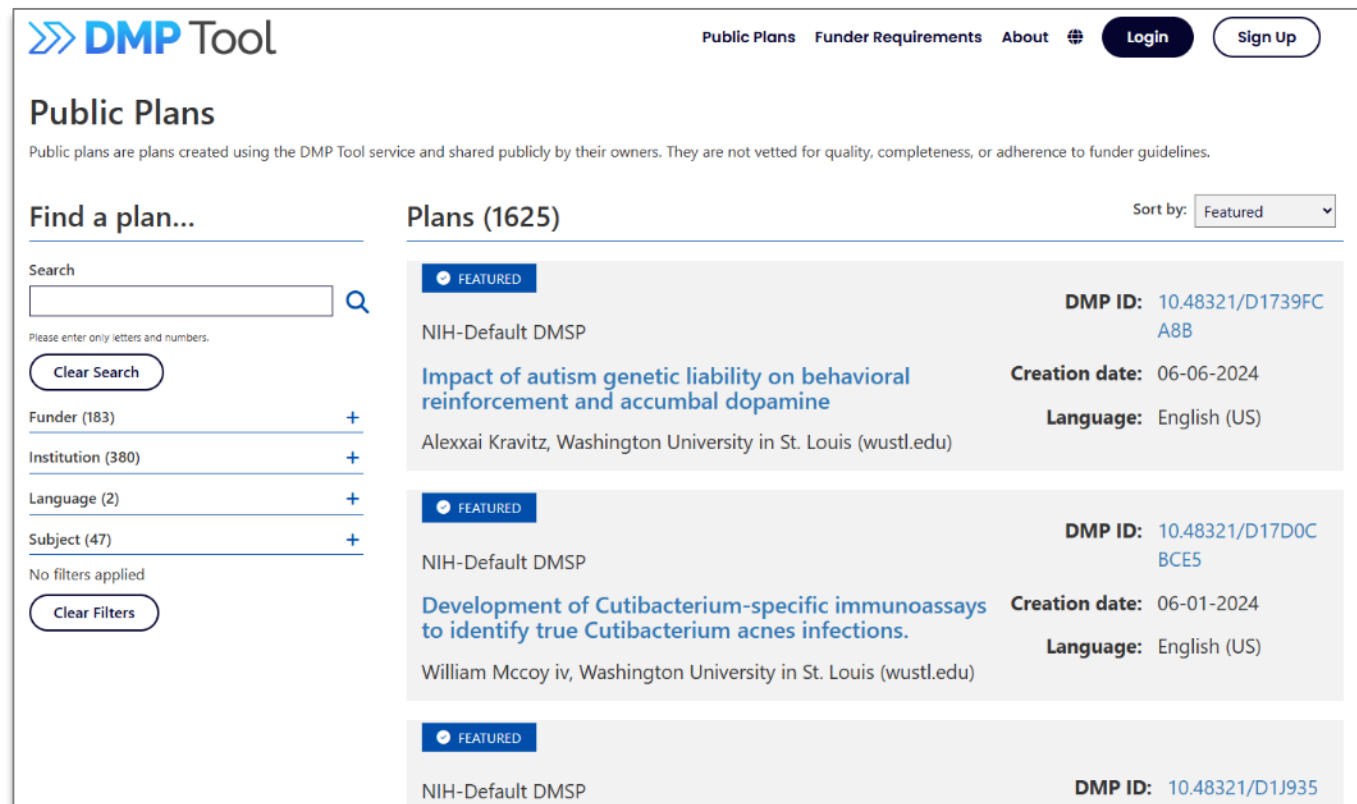
研究開始前に DMP Tool を参考にして DMP を作成

研究データを管理するストレージを Data Storage Finder で選定

研究スタート！

研究中也、研究データのデジタル化や分析までサポート

- 既存の DMP を参考にして作成を進めることができる  
資金提供者・機関・言語・主題で絞り込み、用語検索
- 各資金提供者ごとのテンプレートもあり



The screenshot shows the 'Public Plans' section of the DMP Tool website. The header includes the DMP Tool logo and navigation links for 'Public Plans', 'Funder Requirements', 'About', 'Login', and 'Sign Up'. The main heading is 'Public Plans', followed by a disclaimer: 'Public plans are plans created using the DMP Tool service and shared publicly by their owners. They are not vetted for quality, completeness, or adherence to funder guidelines.'

Below the heading, there is a 'Find a plan...' section on the left with a search bar and filters for Funder (183), Institution (380), Language (2), and Subject (47). A 'Clear Search' button is also present. The main content area displays 'Plans (1625)' with a 'Sort by: Featured' dropdown. Three featured plans are listed:

Plan Title	DMP ID	Creation date	Language
NIH-Default DMSP <b>Impact of autism genetic liability on behavioral reinforcement and accumbal dopamine</b> Alexxai Kravitz, Washington University in St. Louis (wustl.edu)	10.48321/D1739FC A8B	06-06-2024	English (US)
NIH-Default DMSP <b>Development of Cutibacterium-specific immunoassays to identify true Cutibacterium acnes infections.</b> William Mccoy iv, Washington University in St. Louis (wustl.edu)	10.48321/D17D0C BCE5	06-01-2024	English (US)
NIH-Default DMSP	10.48321/D1J935		



# Data Storage Finder



- 研究で扱うデータの特徴を入力するだけで、利用できるストレージサービスを絞り込み比較できる
- コーネル大学図書館職員とコーネル大学IT部門が内製コードが [GitHub](#) で公開されている

The screenshot displays the 'Data Storage Finder' web application. On the left, a 'Describe your data' section asks for data classification. A yellow box highlights the first question: '1. What is the classification of your data?'. The options are: Public / Low Risk, Sensitive / Moderate Risk, Confidential or Restricted / High Risk (selected), and HIPAA-Regulated. The second question is '2. Do you need backups, snapshots or replication of your data?'. The main area shows a grid of storage services with selection buttons. A yellow arrow points from the 'CCSS Data & Reproduction Archive' service to the comparison table on the right.

**Describe your data**

Answer these questions to help identify data storage services that are suitable for your needs. Checking these boxes will change the list of available services. If you are uncertain how to answer, leave the question blank to maximize your resulting options.

Clear Answers

1. What is the classification of your data?

- ☐ Public / Low Risk
- ☐ Sensitive / Moderate Risk
- ☒ Confidential or Restricted / High Risk
- ☐ HIPAA-Regulated

2. Do you need backups, snapshots or replication of your data?

Select data storage services you would like to compare.

Select All Clear Selections

Service	Description
Amazon Web Services Elastic Block Store	Storage for use with Amazon EC2
Amazon Web Services Elastic File System	Storage for use with multiple Amazon EC2 instances
Amazon Web Services Glacier	Cloud based archival storage
Amazon Web Services S3	Flexible, scalable object storage
BioHPC Cloud	Storage for BioHPC lab computing services
CAC Archival Storage	Single copy, non-mountable storage
CAC Red Cloud Storage	Storage for Red Cloud compute instances
CCSS Data & Reproduction Archive	Publicly shared and restricted data and code packages repository
CCSS Regulated Research Environment	Regulated data computing environment
CCSS Research Servers	Storage for CCSS computing environment
CUGIR	Publicly shared geospatial data storage
CUL eCommons	Publicly shared data repository

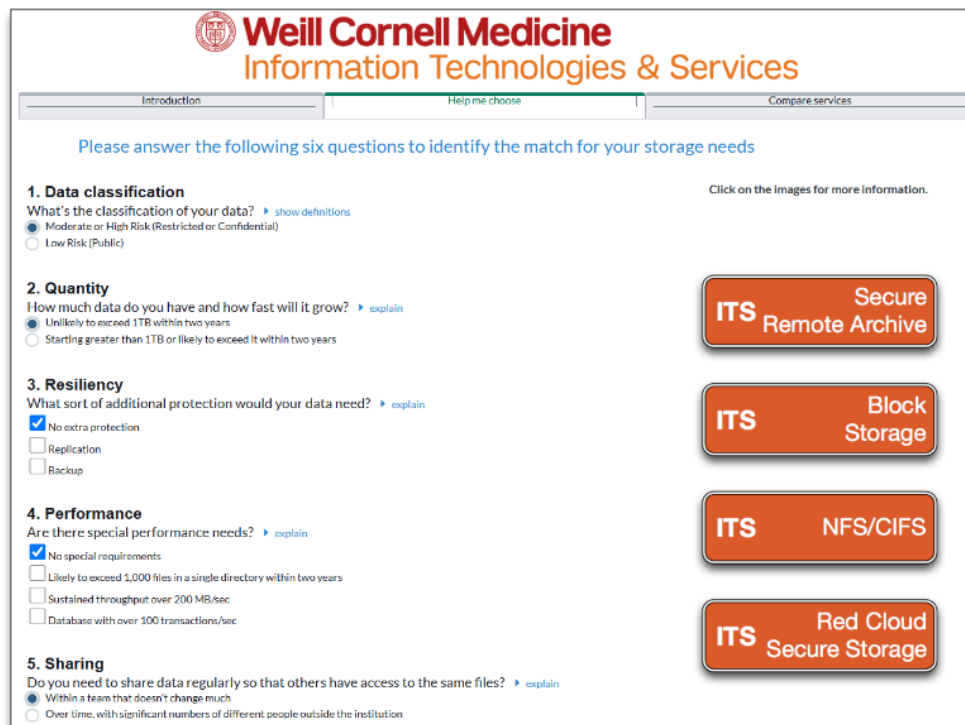
**Compare services that match your selected criteria**

Service	Amazon Web Services Elastic Block Store	Amazon Web Services S3
Brief Description	Amazon Elastic Block Store (EBS) provides persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud. EBS volumes are automatically replicated to provide high availability and consistency. EBS offers high performance and the ability to scale your storage up or down quickly while paying only for what you use.	Amazon S3 is easy to use object storage, with a simple web service interface to store and retrieve any amount of data from anywhere on the web. Pay only for the storage you actually use, with no minimum fee and no setup cost. Amazon S3 offers a range of storage classes designed for different use cases, and configurable policies for managing your data throughout its lifecycle.
Example Use	Storing data files to be accessed by a single Amazon EC2 server.	Making a large dataset available for public download. Private file sharing with versioning and replication, for a widely distributed team.
Cost	Cost dependent on storage class, replication and other services used.	Cost dependent on storage class, replication and other services used.
Capacity	No practical file size limit. No practical overall limit. No practical limit to number of files (costs incurred).	S3 file size limit. No overall limit (costs incurred). No practical limit to number of files.
Access and Collaboration	End user access control to the content of EBS volumes is implemented by the EC2 instance to which an EBS volume is attached. Access and management of EBS volumes themselves is controlled by AWS IAM (Identity and Access Management), which can be linked to Cornell Active Directory for authentication and authorization.	Extensive control using AWS IAM (Identity and Access Management) and Cornell Active Directory users and groups.
Data Allowed	Allowed: Public data. Sensitive / moderate data. Allowed with special configuration: HIPAA-protected data. Confidential or restricted / high risk data. Contact <a href="#">Amazon Web Services Health, Privacy &amp; Compliance</a> for more information. Not Allowed: HIPAA regulated data. Export Controlled information.	Allowed: Public data. Sensitive / moderate risk data. Allowed with special configuration: HIPAA-protected data. Confidential or restricted / high risk data. Contact <a href="#">Amazon Web Services S3</a> for more information. Not Allowed: HIPAA regulated data. Export Controlled information.
Durability Protection	High	High

扱うデータの特徴について  
チェックを入れてください

選択したストレージ  
間で比較ができる

- 限られた管理者のみアクセス可能なセキュアなシステム
- AWS 上に完全に閉じた計算資源を確保し、外部ネットワークとの接続ができないようになっている
- こちら研究データの特性に合わせてストレージを選定可能



**Weill Cornell Medicine**  
Information Technologies & Services

Introduction | Help me choose | Compare services

Please answer the following six questions to identify the match for your storage needs

Click on the Images for more information.

**1. Data classification**  
What's the classification of your data? [show definitions](#)  
☒ Moderate or High Risk (Restricted or Confidential)  
☐ Low Risk (Public)

**2. Quantity**  
How much data do you have and how fast will it grow? [explain](#)  
☒ Unlikely to exceed 1TB within two years  
☐ Starting greater than 1TB or likely to exceed it within two years

**3. Resiliency**  
What sort of additional protection would your data need? [explain](#)  
☒ No extra protection  
☐ Replication  
☐ Backup

**4. Performance**  
Are there special performance needs? [explain](#)  
☒ No special requirements  
☐ Likely to exceed 1,000 files in a single directory within two years  
☐ Sustained throughput over 200 MB/sec  
☐ Database with over 100 transactions/sec

**5. Sharing**  
Do you need to share data regularly so that others have access to the same files? [explain](#)  
☒ Within a team that doesn't change much  
☐ Over time, with significant numbers of different people outside the institution

**ITS Secure Remote Archive**

**ITS Block Storage**

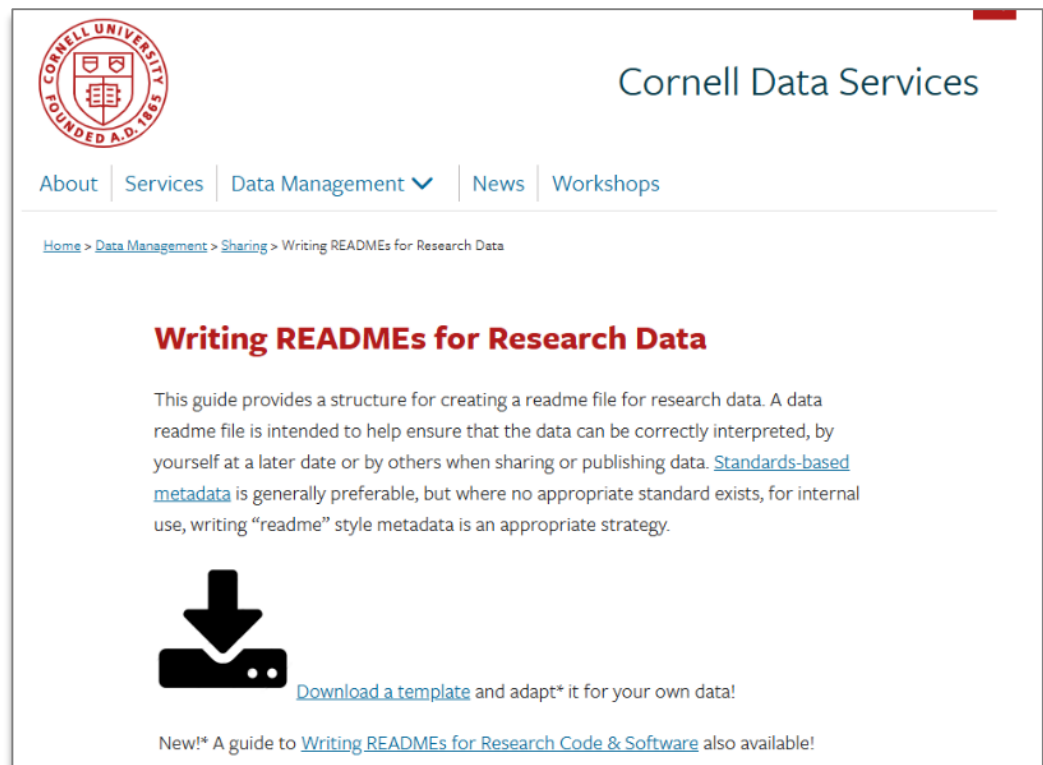
**ITS NFS/CIFS**

**ITS Red Cloud Secure Storage**

- 分野によってスタンダードな形式あり

- readme

連絡先、データに適用される制限やライセンス情報、引用文献、プロトコルや略語の説明、親フォルダ etc.



The screenshot shows the Cornell Data Services website. At the top left is the Cornell University logo. To the right is the text 'Cornell Data Services'. Below this is a navigation bar with links: 'About', 'Services', 'Data Management' (with a dropdown arrow), 'News', and 'Workshops'. Below the navigation bar is a breadcrumb trail: 'Home > Data Management > Sharing > Writing READMEs for Research Data'. The main heading is 'Writing READMEs for Research Data' in red. The text below explains that the guide provides a structure for creating a readme file for research data, intended to help ensure the data can be correctly interpreted. It mentions that 'standards-based metadata' is generally preferable, but for internal use, writing 'readme' style metadata is appropriate. Below the text is a large black download icon (a square with a downward arrow). To the right of the icon is the text 'Download a template and adapt\* it for your own data!'. At the bottom, it says 'New!\* A guide to Writing READMEs for Research Code & Software also available!'.

“Writing READMEs for Research Data”. Cornell Data Services.  
<https://data.research.cornell.edu/data-management/sharing/readme/>  
(最終閲覧日 : 2025/5/26)



データキュレーターチームが事前にデータ確認

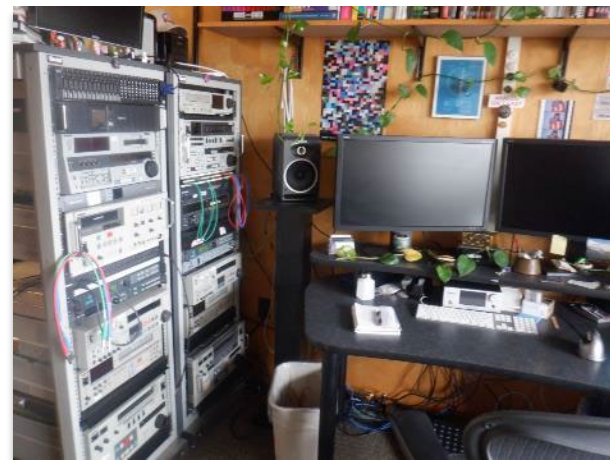
研究者がデータセット、ドキュメント、メタデータファイル、  
readme ファイル等を提供



ファイルレベルでキュレーション、フィードバックとアドバイス  
ファイルが揃っているか、開いて正常に動作するか

- DCN（Data Curation Network）に所属  
データセットを DCN に送信 → 他機関のキュレーターへ

アナログデータからデジタルデータへの変換  
様々な機材を用いて専門スタッフがサポート



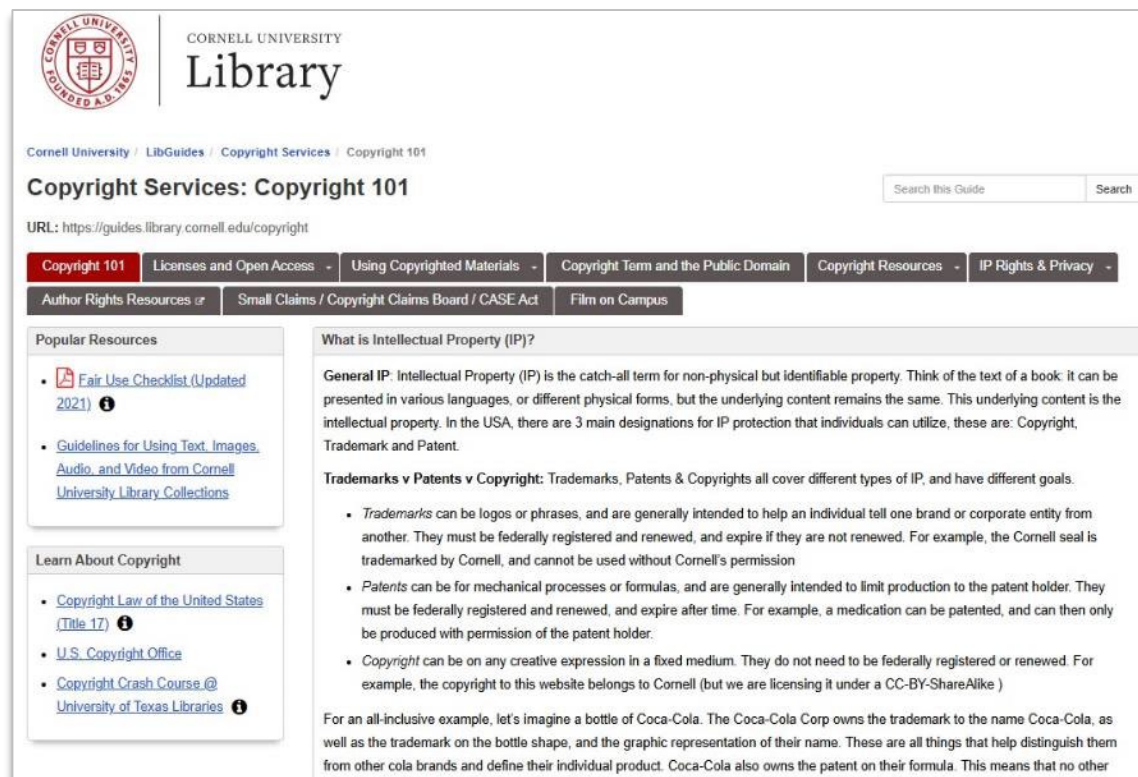
- デジタルデータを扱うための支援  
ライブラリアン3名（+図書館の他スタッフもサポート）

例えば.....

テキストマイニング、機械学習やアルゴリズムの使用、  
データベースの作成、メタデータ付与、データの可視化  
等を支援（かなり専門的）

- 様々な学部 of 学生や研究者の集まる場として、  
関連する研究内容を持つ研究者を紹介するという支援も

- 包括的なウェブサイトを作成  
パブリックドメインのチャートが特に人気
- ウェブサイト+アウトリーチで提供する教育活動



- 過去には著作権サービスチームが構成されており  
図書館の著作権に関すること全般を担当  
現在は教員の出版契約に重点を置く
- 教員は著作権やフェアユースに関する意識があまりない  
出版契約に署名する前に質問してもらい、何に注意すべきか  
知ってもらう（署名した後になって権利を失うことになると思う）
- 博士課程の学生のためのプログラムでも著作権を担当  
出版前に注意すべき点について具体的に説明

- 2012年～ ORCID の Organizational Member に所属  
(創設時からの初期メンバー)
- 新入学の大学院生は2学期の終わりまでにORCID ID の取得義務あり  
→ Affiliation Manager を利用して新入生のリストを ORCID に  
一括インポート・管理





- 図書館で ORCID ID を管理しているのは院生のみ  
研究者はトラッキングしていない
- ORCID ID を使って研究者の出版物を追跡することに行っていない  
学部ごとに独自ルール・ポリシーを所持  
＝大学全体で統一的な方法を使用することができない
- eCommons（機関リポジトリ、DSpace）とは未連携  
DSpace をバージョンアップすると著者 ID と ORCID ID を  
統合可能になる

- OSTP の研究成果公開方針において、パブリックアクセスを実現する方法とその資金については明確にされていない
- 資金調達における大学の役割と責任について、コーネル大学内でもはっきりとした合意は得られていない
- APC は高額すぎるので、Gold OA だけで即時OA義務に対応するのは非現実的
- 様々な分野の論文を、同じ大学の著者が書いたから全て同じリポジトリに保存するというのは意味がない
- 機関リポジトリでアーカイブするならば、それらを見つけられるようにする仕組みが必要



OKAYAMA  
UNIVERSITY



～ご清聴ありがとうございました～