

RDUF公開シンポジウム  
～オープンサイエンスを巡る世界の最新動向～

# データ基盤分野

2017年6月19日 15:35-15:50  
科学技術振興機構 (JST) 別館1階ホール

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  
コンテンツ科学研究系  
オープンサイエンス基盤研究センター

込山 悠介

komiyama@nii.ac.jp

# 日本と欧米の研究データ基盤

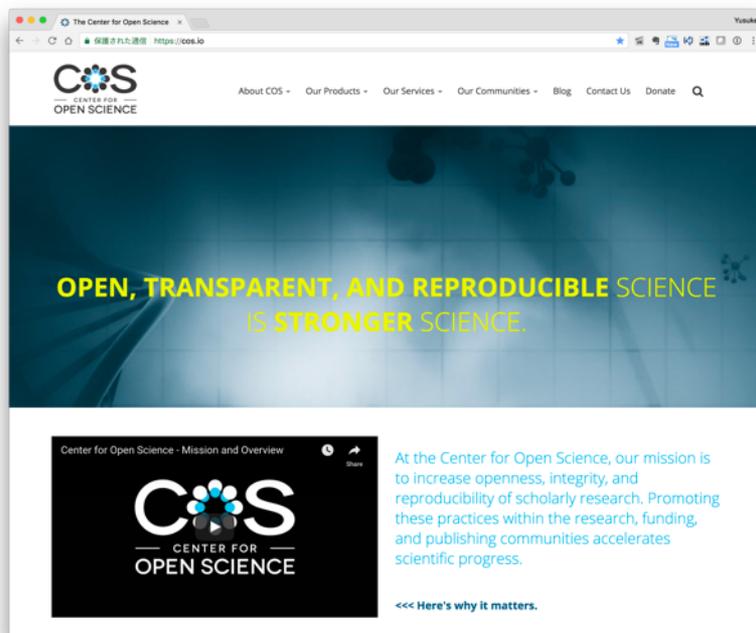
---

- 米国のデータ基盤の事例
- ヨーロッパのデータ基盤の事例
- 日本(国立情報学研究所)で開発中のデータ基盤の紹介

# アメリカのデータ基盤の事例

2017/6/20

# 米国の事例：Center for Open Science



<https://cos.io/>

- 米ヴァージニア州シャーロットビル市にあるNPO。バージニア大学の心理学の研究者Brian Nosek教授らのグループがLaura and John Arnold財団の基金を受け2013年設立。
- 次に紹介するOpen Science FrameworkやSHARE等の研究データ基盤のオープンソースが代表的なプロダクト。
- 心理学発ではあるが、50人近いエンジニアと普及啓蒙のためのスタッフ20人ほど抱えており、盛んにWebiner等が行われている。

B. A. Nosek et al., "SCIENTIFIC STANDARDS. Promoting an open research culture.," Science, vol. 348, no. 6242, pp. 1422–5, Jun. 2015.

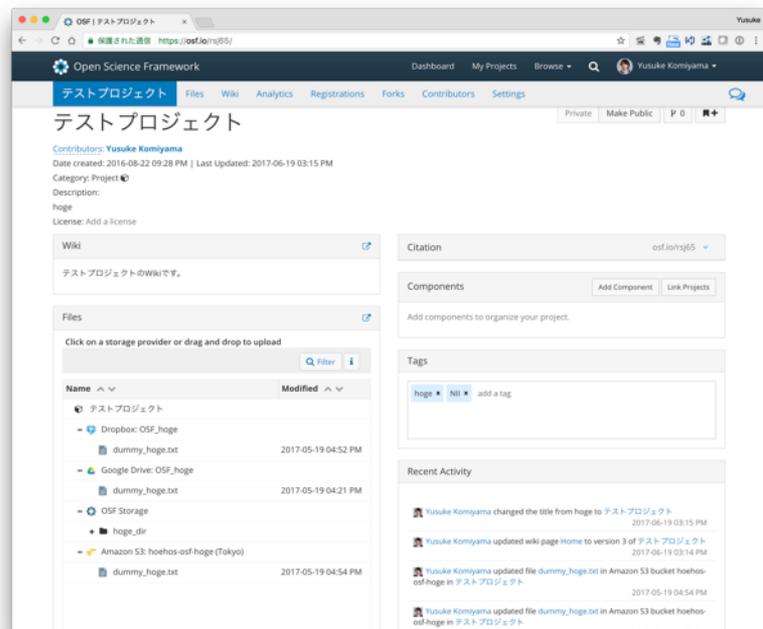
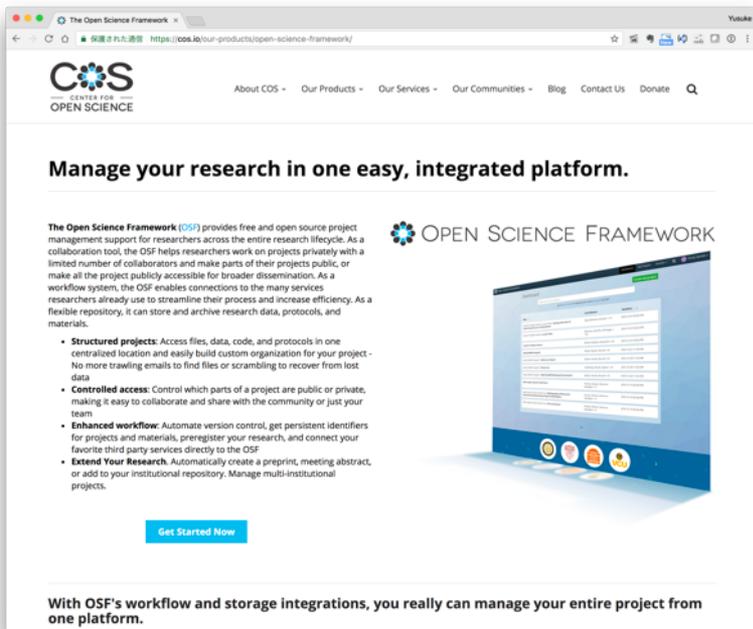
2017/6/20

# COSの研究データ基盤

## Open Science Framework (OSF)

研究データ管理のためのWebサービス (SaaS) 提供およびOSS公開

OSFの研究データ管理機能



E. D. Foster, MSLS and A. Deardorff, MLIS, "Open Science Framework (OSF)," J. Med. Libr. Assoc., vol. 105, no. 2, p. 38, Apr. 2017.

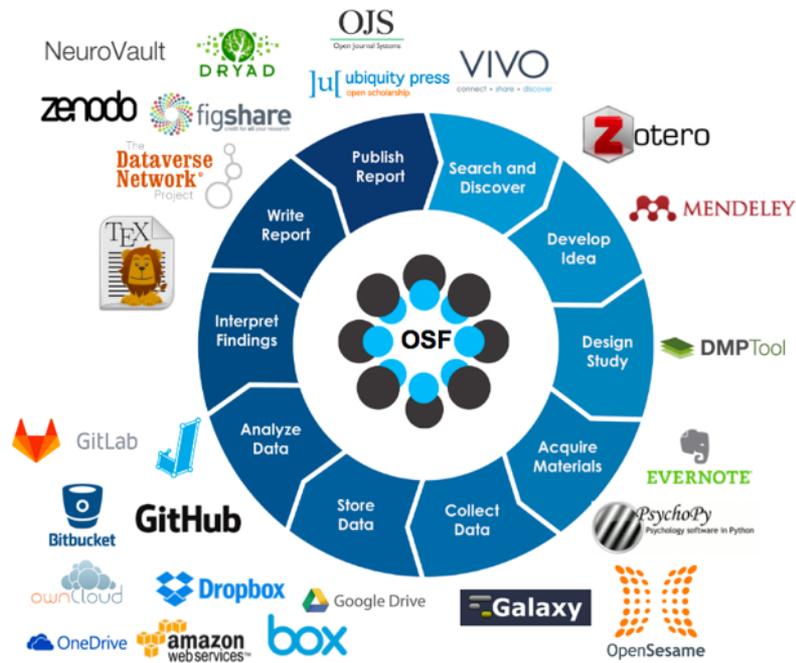
<https://cos.io/our-products/open-science-framework/>

2017/6/20

# COSの研究データ基盤

## Open Science Framework (OSF)

研究に必要なツールやストレージの統合



<https://cos.io/our-products/open-science-framework/>

2017/6/20

SCIENTIFIC DATA (NPG)の recommended data repository

**Social science**

Harvard Dataverse	view re3data entry
UK Data Service ReShare	view re3data entry
openCPSR	view re3data entry
Open Science Framework	view BioSharing entry
Archaeology Data Service	view re3data entry

**Generalist repositories**

Scientific Data encourages authors to archive data to one of the above data-type specific repositories where possible. Where a data-type specific repository is not available, we recommend the following generalist repositories, which can handle a wide variety of data. Generalist repositories may also be appropriate for archiving associated analyses, or experimental-control data, supplementing the primary data in a data-type specific repository.

Repository Name	Information on fees/costs	Size limits	Integrated with Scientific Data's manuscript submission system	Re3data / BioSharing entry
Dryad Digital Repository	\$120 USD for first 20 GB, and \$50 USD for each additional 10 GB	None stated	Yes ✓	view BioSharing entry
figshare	100 GB free per Scientific Data manuscript. Additional fees apply for larger datasets	1 TB per dataset	Yes ✓ - To qualify for the 100 GB of free storage, data must be uploaded to figshare via our submission system. Download instructions.	view BioSharing entry
Harvard Dataverse	Contact repository for datasets over 1 TB	2.5 GB per file, 10 GB per dataset	No	view re3data entry

<https://www.nature.com/sdata/policies/repositories>

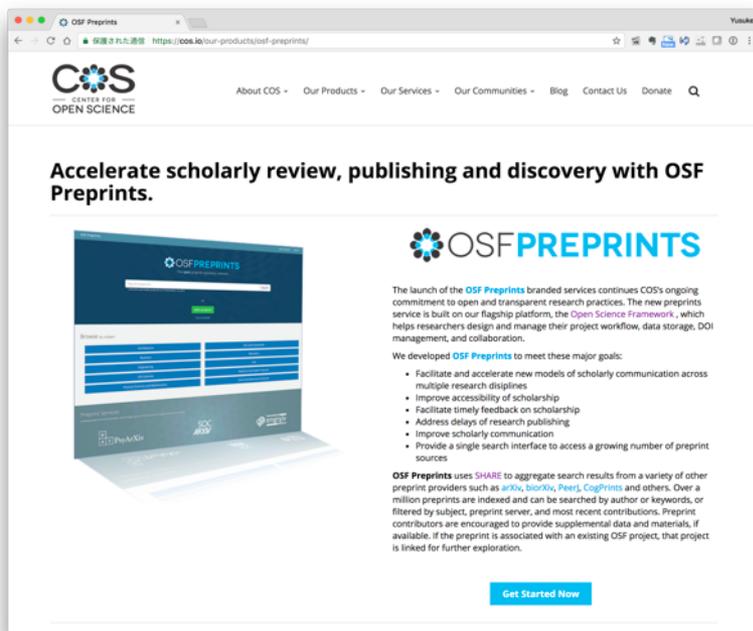
# OSFを用いたプレプリントの展開

## OSF PREPRINTS

### プレプリントのセントラルサービス

AgriXiv, engrXiv, LawArXiv, PsyArXiv, SOC ARXIV

農学、工学、法学、心理学、社会科学などのプレプリントサービスが立ち上がり、Google Scholarで検索できる



OSF PREPRINTS

Accelerate scholarly review, publishing and discovery with OSF Preprints.

**OSF PREPRINTS**

The launch of the **OSF Preprints** branded services continues COS's ongoing commitment to open and transparent research practices. The new preprints service is built on our flagship platform, the **Open Science Framework**, which helps researchers design and manage their project workflow, data storage, DOI management, and collaboration.

We developed **OSF Preprints** to meet these major goals:

- Facilitate and accelerate new models of scholarly communication across multiple research disciplines
- Improve accessibility of scholarship
- Facilitate timely feedback on scholarship
- Address delays of research publishing
- Improve scholarly communication
- Provide a single search interface to access a growing number of preprint sources

**OSF Preprints** uses **SHARE** to aggregate search results from a variety of other preprint providers such as **arXiv, bioRxiv, PeerJ, CogPrints** and others. Over a million preprints are indexed and can be searched by author or keywords, or filtered by subject, preprint server, and most recent contributions. Preprint contributors are encouraged to provide supplemental data and materials, if available. If the preprint is associated with an existing OSF project, that project is linked for further exploration.

[Get Started Now](#)

### You build the community. We've built the infrastructure.

COS provides branded preprint services to communities interested in building a preprint service for a specific discipline, research topic, geography or other commonality. The branded preprint services are built on the **OSF Preprints** infrastructure, a shared toolset for file upload, sharing, persistence, and discovery.

**OSF Preprints** and branded preprint services are integrated with **SHARE** for search and discovery. **SHARE** allows users to search across all preprint platforms, including those hosted by other groups such as **arXiv, bioRxiv, and PeerJ**. You can focus your efforts on generating research, not on managing research infrastructure.

The service can be configured to match each organization's brand, editorial focus, licensing requirements, and taxonomy. Once configured, a new service can be up and indexing new preprints in a matter of days. We have already launched five major new branded services—**PsyArXiv, AgriXiv, SocArXiv, engrXiv** and **LawArXiv**—and are adding more all the time. These partnerships provide free, open access, open source archives for psychology, social, and engineering science research and are in response to the growing need for faster, open sharing of research on open platforms.



[Get Started Today](#)



All OSF Preprints and their branded affiliates are indexed in Google Scholar for maximum discoverability.

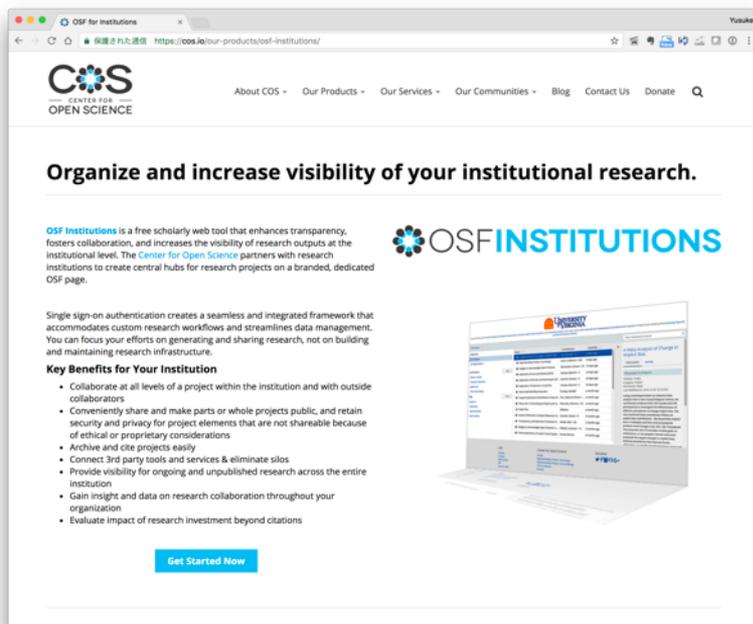
<https://cos.io/our-products/osf-preprints/>

# COSのその他のサービス

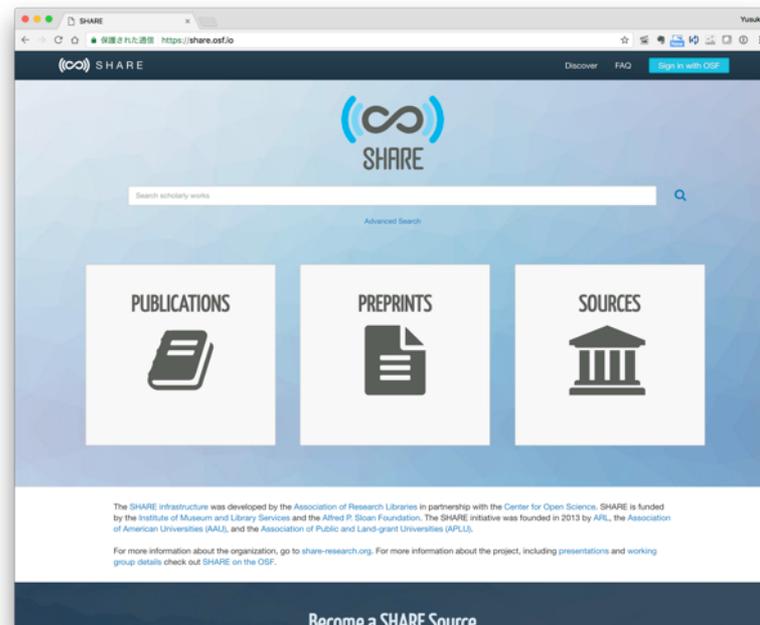
## OSF INSTITUTIONS

OSFベースの機関レポジトリ

SHARE: Association of Research Libraries (米) とCOSの連携で開発されたメタデータの統合検索



<https://cos.io/our-products/osf-institutions/>



<https://share.osf.io/>

2017/6/20

# 米国の大学のデータ基盤事例

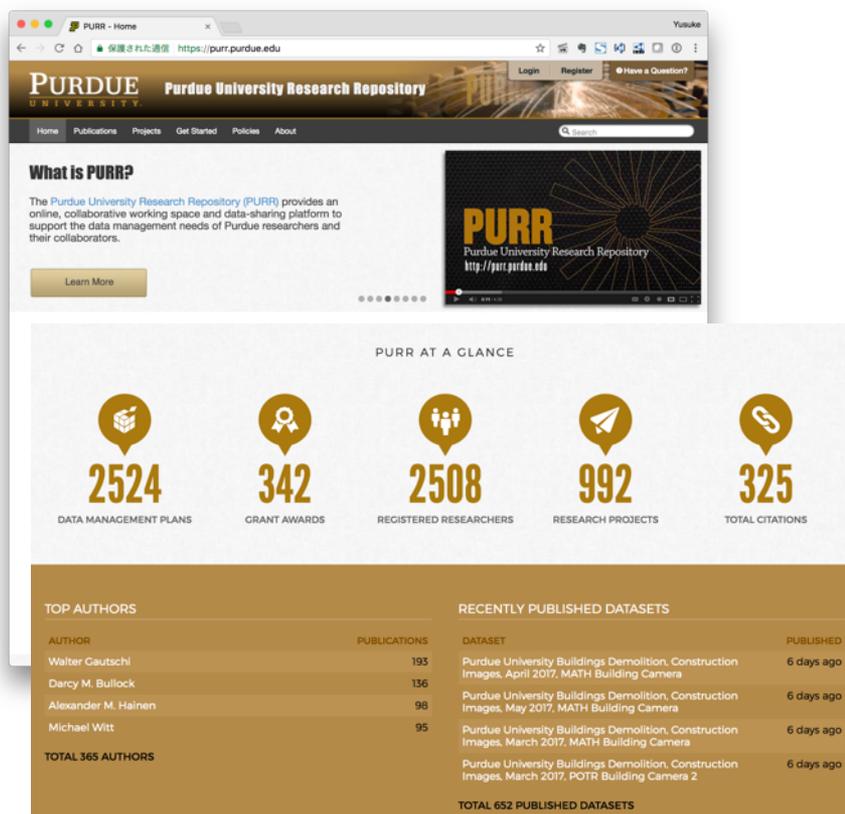
PURR: Purdue University Research Repository

図書館系

globus

e-Science系

(The University of Chicago and Argonne National Laboratory)



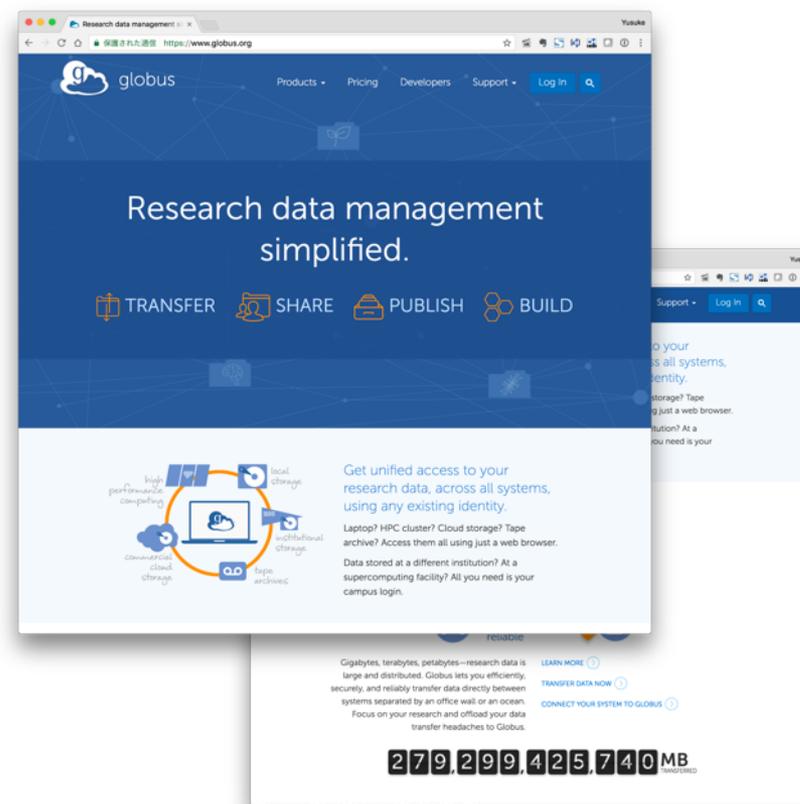
**PURR**  
Purdue University Research Repository  
<http://purrr.purdue.edu>

**What is PURR?**  
The Purdue University Research Repository (PURR) provides an online, collaborative working space and data-sharing platform to support the data management needs of Purdue researchers and their collaborators.

**PURR AT A GLANCE**

- 2524 DATA MANAGEMENT PLANS
- 342 GRANT AWARDS
- 2508 REGISTERED RESEARCHERS
- 992 RESEARCH PROJECTS
- 325 TOTAL CITATIONS

TOP AUTHORS		RECENTLY PUBLISHED DATASETS	
AUTHOR	PUBLICATIONS	DATASET	PUBLISHED
Walter Gautschi	193	Purdue University Buildings Demolition, Construction Images, April 2017, MATH Building Camera	6 days ago
Darcy M. Bullock	136	Purdue University Buildings Demolition, Construction Images, May 2017, MATH Building Camera	6 days ago
Alexander M. Hainan	98	Purdue University Buildings Demolition, Construction Images, March 2017, MATH Building Camera	6 days ago
Michael Witt	95	Purdue University Buildings Demolition, Construction Images, March 2017, POTR Building Camera 2	6 days ago
<b>TOTAL 365 AUTHORS</b>		<b>TOTAL 652 PUBLISHED DATASETS</b>	



**globus**

Research data management simplified.

TRANSFER SHARE PUBLISH BUILD

Get unified access to your research data, across all systems, using any existing identity.

Laptop? HPC cluster? Cloud storage? Tape archive? Access them all using just a web browser.

Data stored at a different institution? At a supercomputing facility? All you need is your campus login.

279,299,425,740 MB

C. C. Dearborn, et al., OCLC Syst. Serv. Int. Digit. Libr. Perspect., vol. 30, no. 1, pp. 15–27, Feb. 2014.

2017/6/20

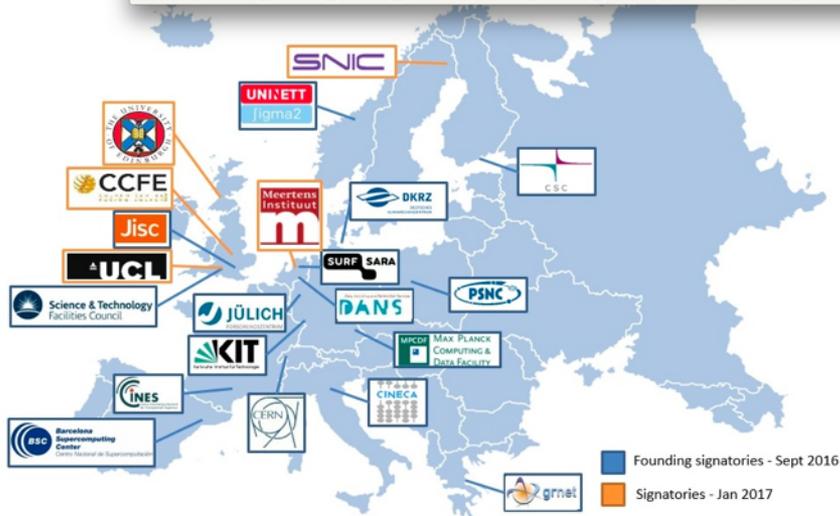
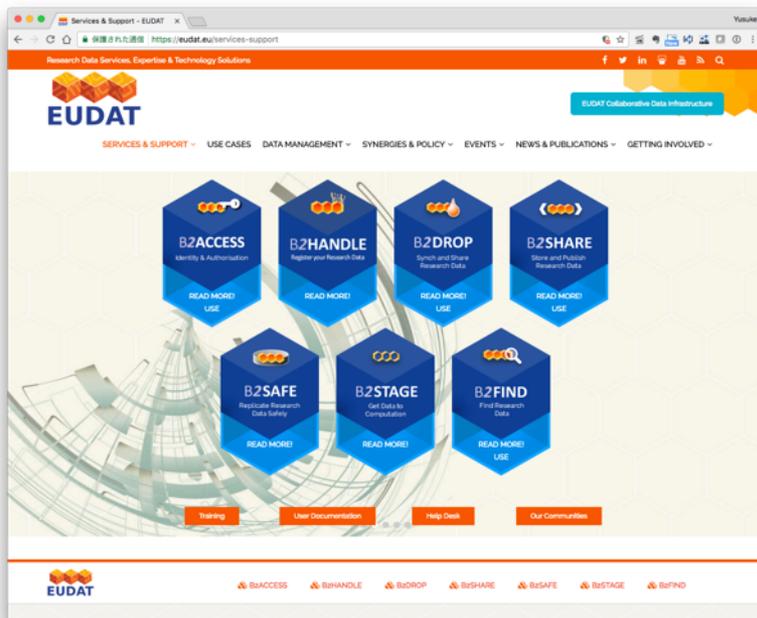
<https://purrr.purdue.edu/>

R. Ananthkrishnan, et al. "Concurr. Comput.", vol. 27, no. 2, pp. 290–305, Feb. 2015.

<https://www.globus.org/>

# ヨーロッパのデータ基盤の事例

# ヨーロッパの事例: EUDAT



<https://www.eudat.eu/>

- ヨーロッパの研究コミュニティで国境や分野を越えてデータ共有・利活用するためのインフラサービス。EUDAT CDI (Collaborative Data Infrastructure) 参加機関で構成される。
- EUDATのサービス群
  - ストレージ
  - レポジトリ管理
  - メタデータ検索
  - データ管理ポリシーと複製
  - HPC対応
  - 認証連携
  - Persistent ID

D. Lecarpentier et al., Int. J. Digit. Curation, vol. 8, no. 1, pp. 279–287, Jun. 2013.

S. B. Ardestani et al., in 2015 IEEE 11th International Conference on e-Science, 2015, pp. 448–453.

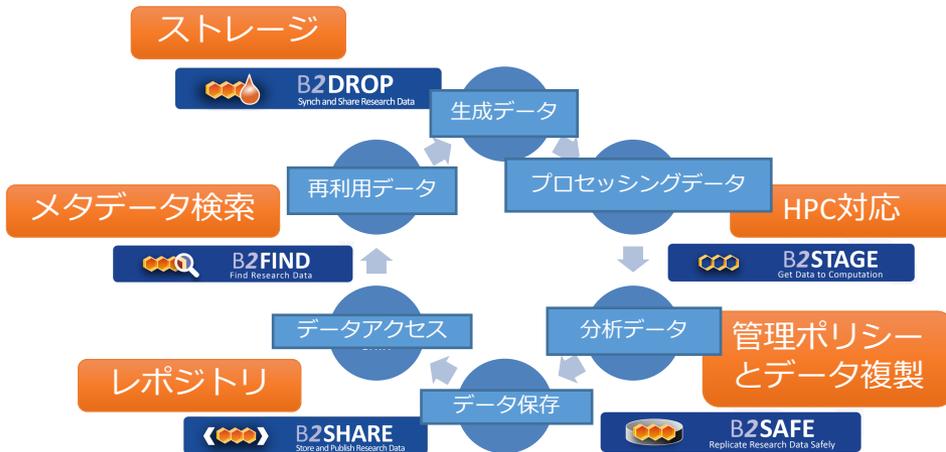
A. S. Memon, et al., in 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing, 2014, pp. 726–731.

# EUDATサービス

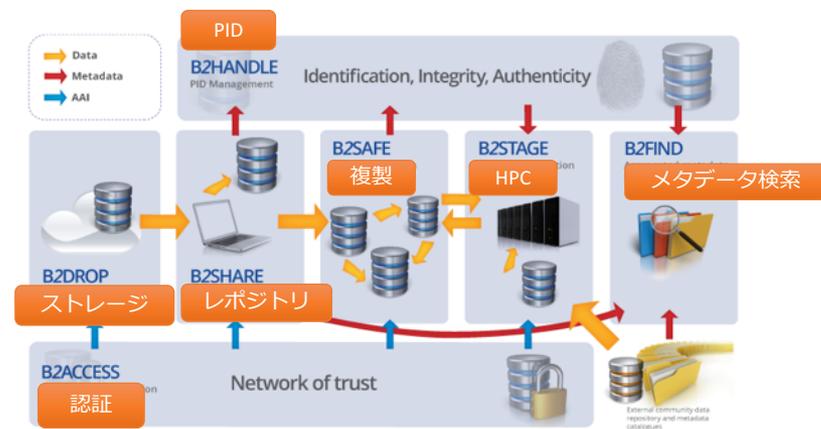
研究におけるデータ利用のサイクルと  
初期のEUDATAサービス



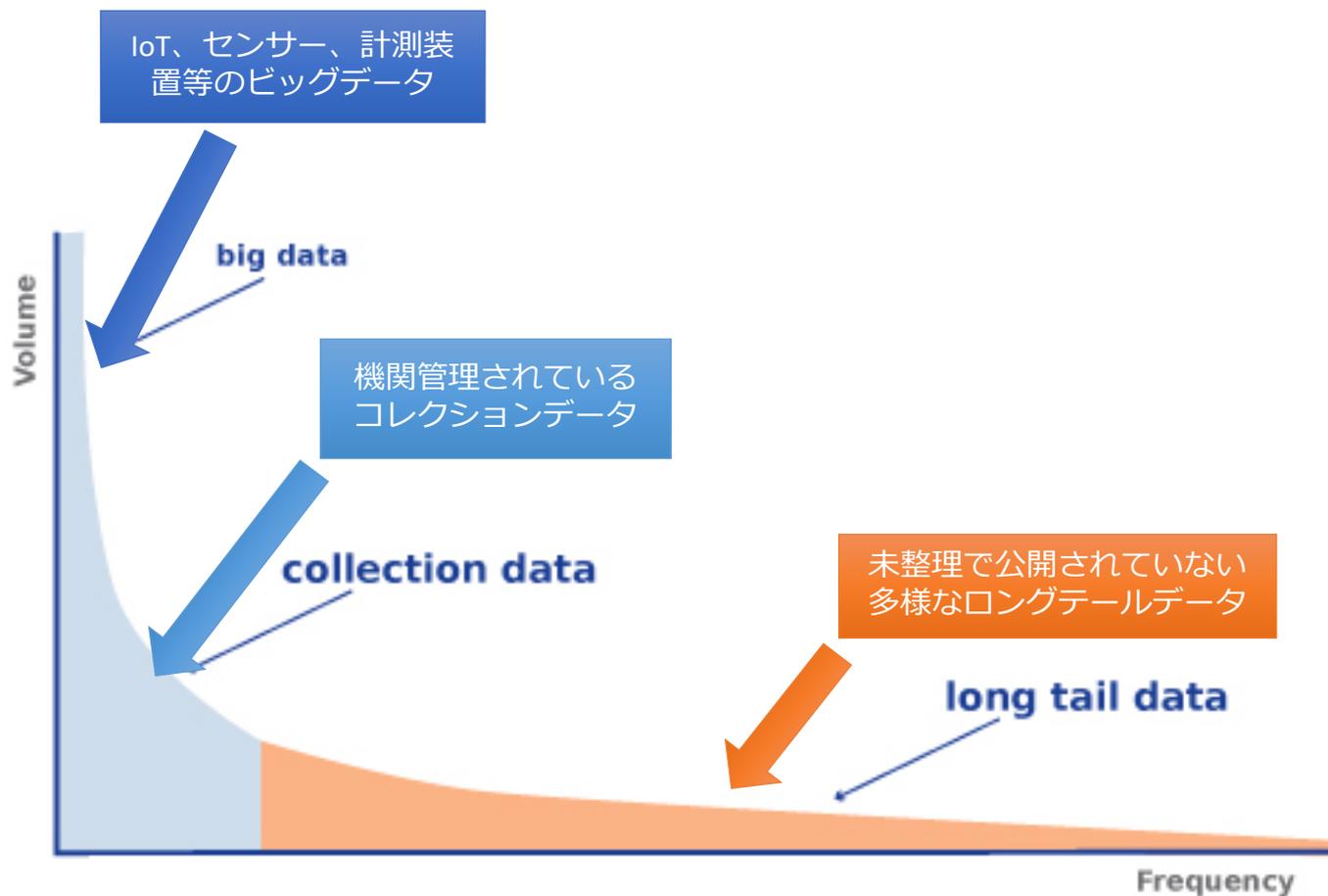
Support throughout the lifecycle



EUDATサービスにおけるデータの流れ  
(B2HANDLEとB2ACCESSは後から追加された)



# ビッグデータとロングテールデータ



<https://www.eudat.eu/eudat-service-road-map>

2017/6/20

# 日本(国立情報学研究所)で 開発中のデータ基盤の事例

## 大学などの学術研究・教育活動の連携・推進



### 学術情報の公開・共有

- ◆ 学術情報流通と オープンアクセスの推進
- ◆ 大学の 機関リポジトリ 拡充の推進



学術コンテンツ基盤

### 大学間連携支援

- ◆ 仕様統一したシステムによる 大学間連携、各種資源の相互利用の促進



### クラウド活用支援

- ◆ クラウド利活用促進による 大幅なIT経費削減・研究教育環境の高度化



クラウド支援サービス

SINET直結クラウド

### セキュリティ強化

- ◆ 暗号技術活用による 情報の保護、安全な認証
- ◆ ネットワーク機能連携による サイバー攻撃対策



### 学術情報ネットワークの構築・運用

- ◆ 国内回線 全国100Gbps化
- ◆ 海外 (米国・欧州・アジア) との高速接続
- ◆ 多様化するニーズに応えるSDNなどの 最新ネットワーク技術の導入



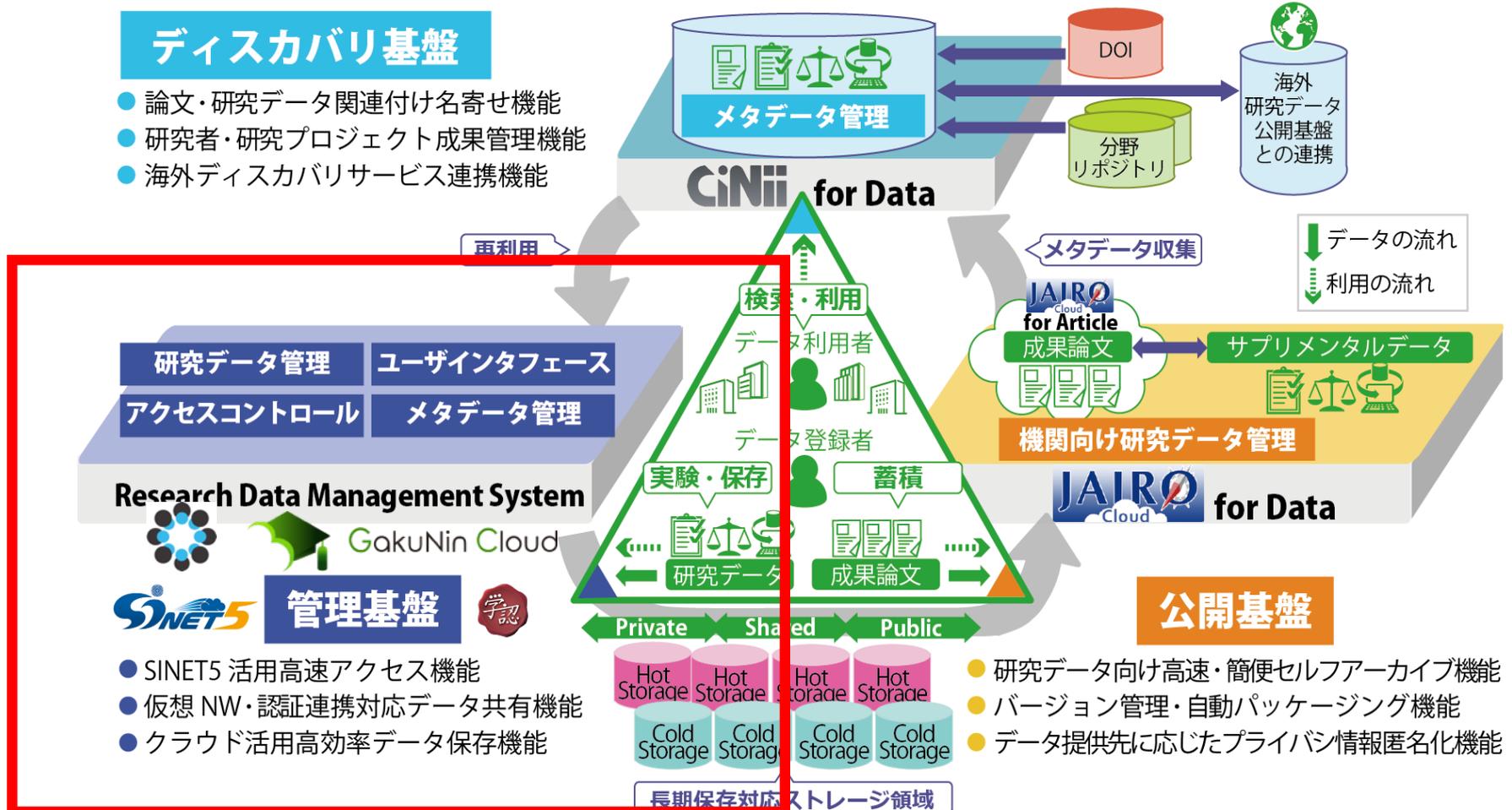
# オープンサイエンス推進のための研究データ基盤

目的

学術コミュニティとの密な連携のもとで最先端の研究データの管理・公開のための共通基盤を整備し、様々な分野でオープンサイエンスの推進を図る

## ディスカバリ基盤

- 論文・研究データ関連付け名寄せ機能
- 研究者・研究プロジェクト成果管理機能
- 海外ディスカバリサービス連携機能



- SINET5 活用高速アクセス機能
- 仮想NW・認証連携対応データ共有機能
- クラウド活用高効率データ保存機能

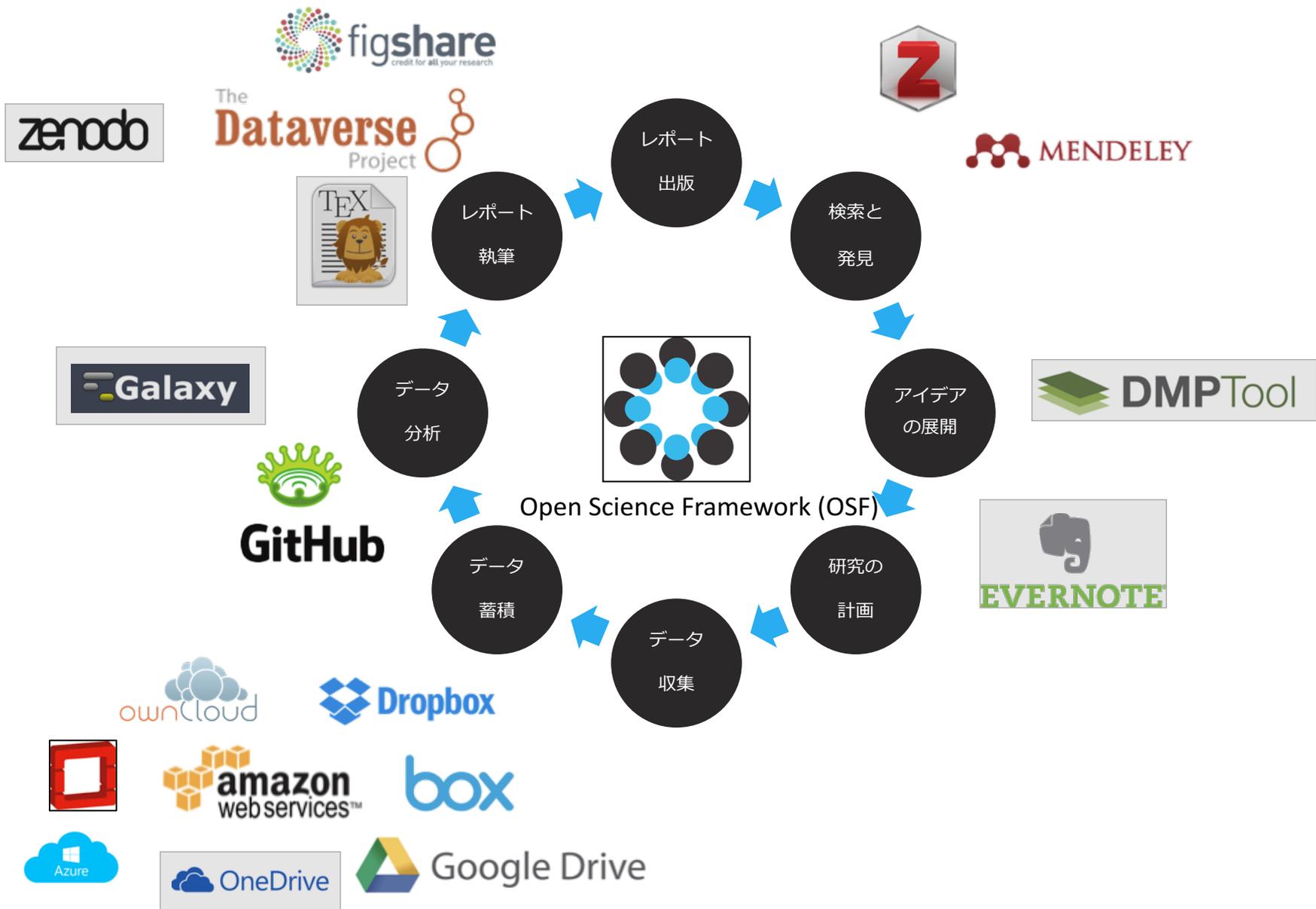
- 研究データ向け高速・簡便セルフアーカイブ機能
- バージョン管理・自動パッケージング機能
- データ提供先に応じたプライバシー情報匿名化機能

「オープンサイエンスイノベーションに資するオープンサイエンスのあり方に関する提言」,  
日本学術会議オープンサイエンスの取組に関する検討委員会より

# 研究目的

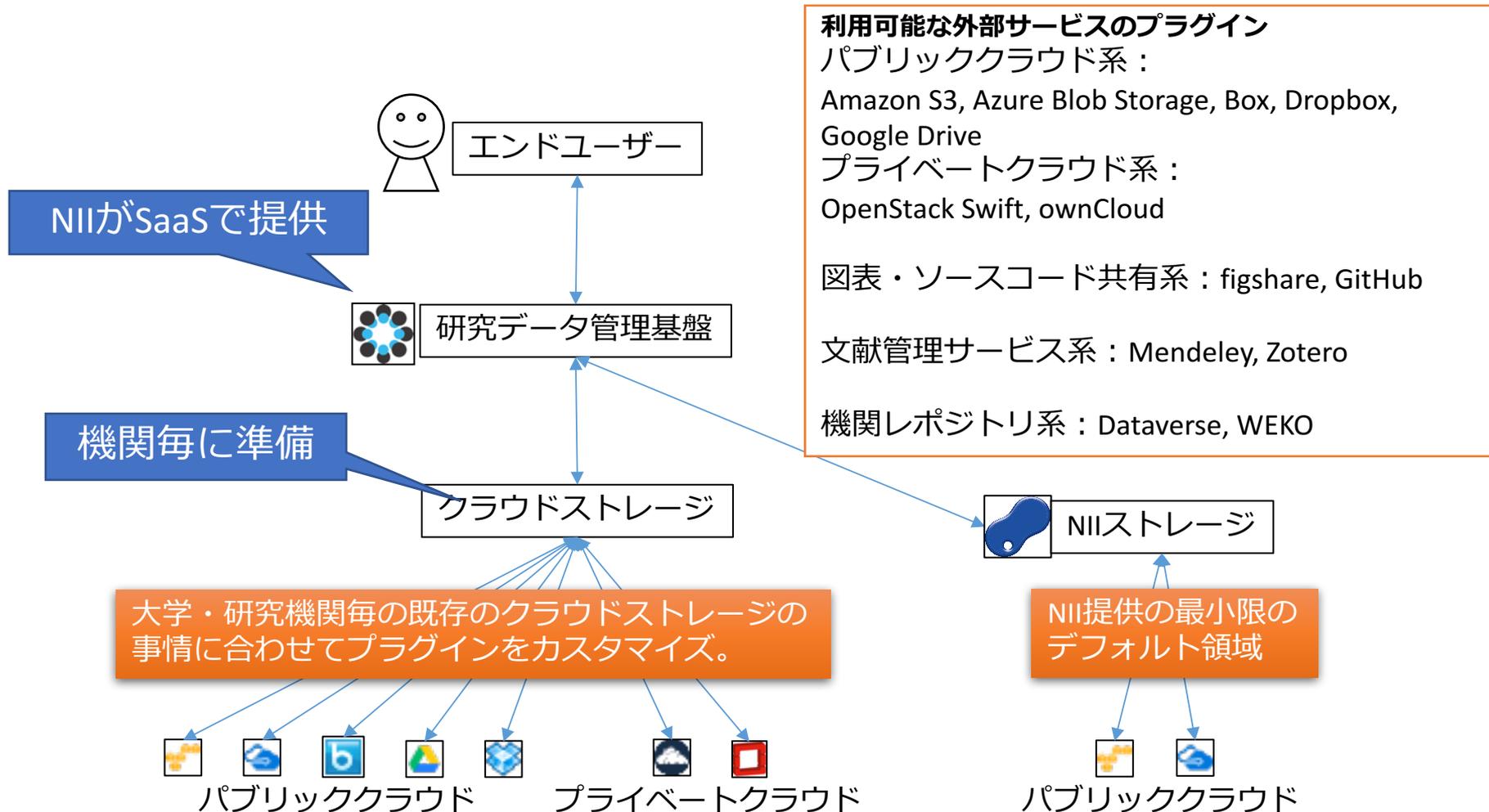
- NIIデータ管理基盤サービスの開発・運営
  - 国立情報学研究所では、オープンソースソフトウェアの **Open Science Framework (OSF)** をベースとし、国内版の**研究データ管理基盤**の発展と成熟を目指す。
  - 研究データ管理基盤サービスをハブとして、研究者が日常的に利用する研究ツールとの連携を充実させる。
  - 機関をまたいだ、共同研究プロジェクトの作業プラットフォームとして、非公開（クローズド）な環境としてサービスを提供。
- NIIオープンサイエンス三基盤との有機的な統合
  - NIIの既存サービス（SINET、学認）や、各大学・研究機関が提供しているクラウドストレージ（パブリック、プライベート）を活用した設計で高いセキュリティ、性能、利便性と管理コストのバランスを考慮する。

# 研究者が日常的に使う研究ツールのハブ



“Opening Science by Opening Workflow”, by Jeff Spies: <https://cos.io/pr/2015-09-24/> より

# NII研究データ管理基盤へのクラウドサービスからの接続



# NII研究データ管理基盤サービスの認証

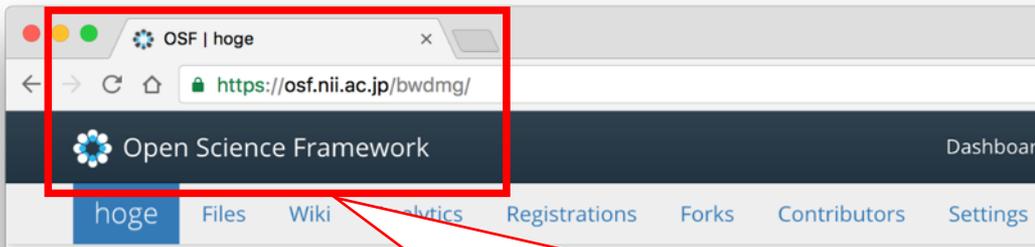
The screenshot shows a web browser window with the URL <https://www.test.rdm.at.nii.ac.jp>. The page title is "研究データ管理基盤 (TEST)" and it features a search bar, a "Support" link, and a "Login" button. A dropdown menu is open, listing various institutions and organizations. The list includes:

- Asahikawa Medical University Test IdP
- Kitami Institute of Technology Test IdP
- Kushiro National College of Technology
- Sapporo Medical University
- ChitoseSamlResearchGroup
- Asahikawa National College of Thechnology
- Hokkaido University
- Muroran Institute of Technology
- sac-sv001.nsrq.fml.org
- Hokkaido Information University
- Yamagata University
- Yamagata University
- Sendai National College of Technology
- Fukushima National College of Technology
- Sendai National College of Technology, Hirose
- Hachinohe National College of Technology
- Akita National College of Technology
- Sendai National College of Technology

The page also displays the text "研究データ管理" and "A scholarly commons to connect", along with a "Hello Hello" message. The footer contains the copyright notice: "Copyright © 2016 National Institute of Informatics".

GakuNin DS

Open Science Frameworkを学認へ対応



## OSFを国内向けに再開発

- 日本国内でNIIドメインでのサービス運営
- UPKI SSLサーバー証明書を利用

## NIIで開発したプラグイン

- NII ストレージ (デフォルト領域)
- Azure Blob Storage アドオン
- Open Stack Swift アドオン
- WEKO (JAIRO Cloud) アドオン



WARN

2017/6/20

# NII研究データ管理基盤サービスの機能紹介

Open Science Framework

Dashboard My Projects Browse

Yusuke Komiyama

My Projects Browse and organize all your projects

Create Project

All my projects

Filter displayed projects

Collections	Name	Contributors	Modified
All my projects	hoge	Komiyama	4 days ago
All my registrations	OSF	Komiyama	11 days ago
Bookmarks (0)	yazawa-test	Yazawa, Komiyama	16 days ago
Contributors	hoge2	Komiyama	16 days ago
Satoshi Yazawa	Fork of Fork of hoge	Komiyama	17 days ago
Tags	Fork of hoge	Komiyama	17 days ago
nii			
osf			
hoge			
国立情報学研究所			

Information Activity

Visibility: Public  
Category: Project  
Permission: Admin  
Last Modified on: 2016-11-09 02:42 PM

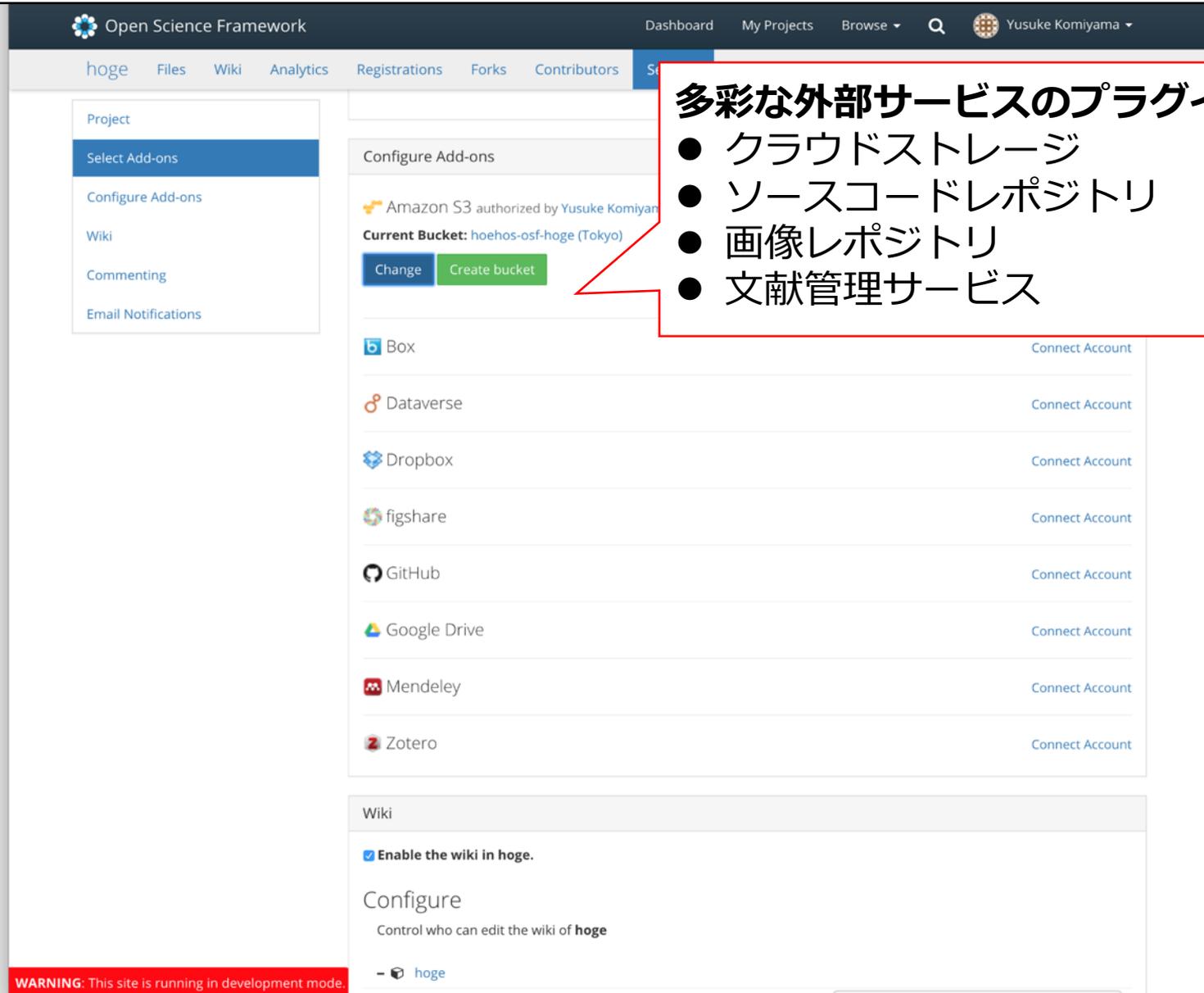
Tags: nii osf hoge 国立情報学研究所

**プロジェクト毎の研究データ管理**

- 共同研究者間での限定共有
- プロジェクトの複製
- プロジェクト毎にクラウドストレージのディレクトリの割当てが可能

WARNING: This site is running in development mode.

Center for Open Science Socialize



Open Science Framework

Dashboard My Projects Browse Yusuke Komiyama

hoge Files Wiki Analytics Registrations Forks Contributors

Project

Select Add-ons

Configure Add-ons

Wiki

Commenting

Email Notifications

Configure Add-ons

Amazon S3 authorized by Yusuke Komiyama

Current Bucket: hoehos-osf-hoge (Tokyo)

Change Create bucket

Box Connect Account

Dataverse Connect Account

Dropbox Connect Account

figshare Connect Account

GitHub Connect Account

Google Drive Connect Account

Mendeley Connect Account

Zotero Connect Account

Wiki

Enable the wiki in hoge.

Configure

Control who can edit the wiki of hoge

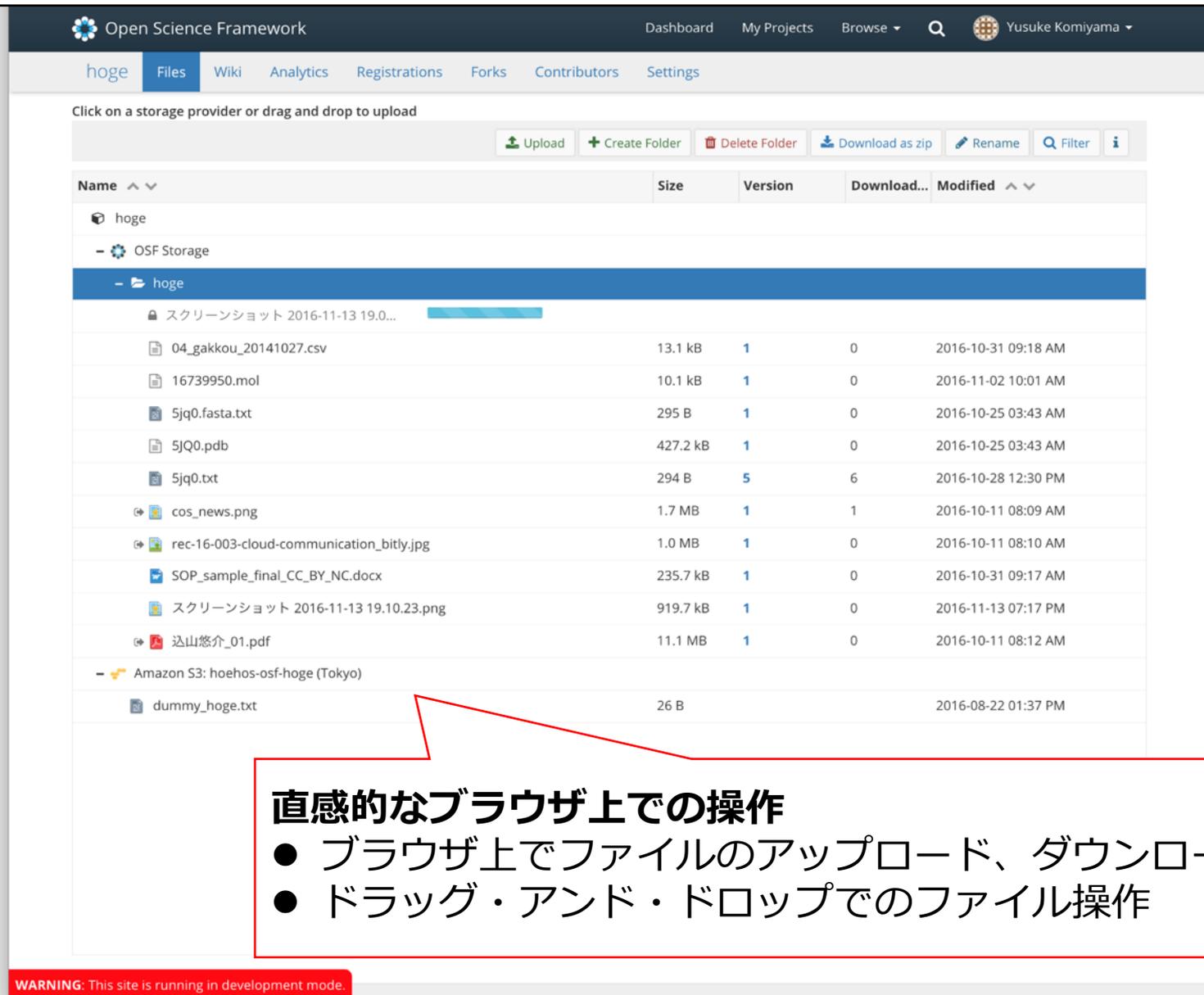
- hoge

WARNING: This site is running in development mode.

## 多彩な外部サービスのプラグイン

- クラウドストレージ
- ソースコードレポジトリ
- 画像レポジトリ
- 文献管理サービス

# NII研究データ管理基盤サービスの機能紹介



The screenshot displays the OSF interface for a project named 'hoge'. The top navigation bar includes 'Open Science Framework', 'Dashboard', 'My Projects', 'Browse', a search icon, and the user name 'Yusuke Komiyama'. Below the navigation bar, there are tabs for 'Files', 'Wiki', 'Analytics', 'Registrations', 'Forks', 'Contributors', and 'Settings'. The main content area shows a file management interface with a toolbar containing 'Upload', 'Create Folder', 'Delete Folder', 'Download as zip', 'Rename', 'Filter', and 'Info' icons. A table lists files and folders, including a folder named 'OSF Storage' and a sub-folder 'hoge'. The 'hoge' folder contains several files such as 'スクリーンショット 2016-11-13 19.0...', '04\_gakkou\_20141027.csv', '16739950.mol', '5jq0.fasta.txt', '5JQ0.pdb', '5jq0.txt', 'cos\_news.png', 'rec-16-003-cloud-communication\_bitly.jpg', 'SOP\_sample\_final\_CC\_BY\_NC.docx', 'スクリーンショット 2016-11-13 19.10.23.png', and '込山悠介\_01.pdf'. Below the 'hoge' folder, there is an 'Amazon S3: hoehos-osf-hoge (Tokyo)' folder containing a file named 'dummy\_hoge.txt'. A red callout box points to the 'dummy\_hoge.txt' file.

Name	Size	Version	Download...	Modified
hoge				
OSF Storage				
hoge				
スクリーンショット 2016-11-13 19.0...				
04_gakkou_20141027.csv	13.1 kB	1	0	2016-10-31 09:18 AM
16739950.mol	10.1 kB	1	0	2016-11-02 10:01 AM
5jq0.fasta.txt	295 B	1	0	2016-10-25 03:43 AM
5JQ0.pdb	427.2 kB	1	0	2016-10-25 03:43 AM
5jq0.txt	294 B	5	6	2016-10-28 12:30 PM
cos_news.png	1.7 MB	1	1	2016-10-11 08:09 AM
rec-16-003-cloud-communication_bitly.jpg	1.0 MB	1	0	2016-10-11 08:10 AM
SOP_sample_final_CC_BY_NC.docx	235.7 kB	1	0	2016-10-31 09:17 AM
スクリーンショット 2016-11-13 19.10.23.png	919.7 kB	1	0	2016-11-13 07:17 PM
込山悠介_01.pdf	11.1 MB	1	0	2016-10-11 08:12 AM
Amazon S3: hoehos-osf-hoge (Tokyo)				
dummy_hoge.txt	26 B			2016-08-22 01:37 PM

## 直感的なブラウザ上での操作

- ブラウザ上でファイルのアップロード、ダウンロード
- ドラッグ・アンド・ドロップでのファイル操作

WARNING: This site is running in development mode.

2017/6/20

# NII研究データ管理基盤サービスの機能紹介

hoge Files Wiki Analytics Registrations Forks Contributors Settings

04\_gakkou\_20141027.csv (Version: 1)

Delete Check out Share Download Toggle view: View Edit Revisions

OSF Storage

- hoge
  - 04\_gakkou\_20141027.csv
  - 16739950.mol
  - 5jq0.fasta.txt
  - 5JQ0.pdb
  - 5jq0.txt
  - cos\_news.png
  - rec-16-003-cloud-commu...
  - SOP\_sample\_final\_CC\_BY...
  - 込山悠介\_01.pdf
- Amazon S3: hoehos-osf-hoge (T...
  - dummy\_hoge.txt

Sheet\_1

Show rows with cells including:

施設名	郵便番号
環境情報センター	〒252-0236
相模川ビレッジ若あゆ...	〒252-0135
城山学校給食センター	〒252-0111
青少年学習センター	〒252-0207
青少年相談センター	〒252-0239
清新学校給食センター	〒252-0217
総合学習センター	〒252-0239
津久井学校給食センター	〒252-0153
津久井生涯学習センター	〒252-0159
ふじの体験の森やませ...	〒252-0182
相原小学校	〒252-0141
相原中学校	〒252-0143
青根小学校	〒252-0162
青根中学校	〒252-0162
青野原小学校	〒252-0161
青野原中学校	〒252-0161
青葉小学校	〒252-0228
旭小学校	〒252-0143

ording to Step-1a or 1b.

cells. Avoid excessive amount.)

l at 300 × g in a tube for 5 minutes to form cell pellet. Carefully proceed to Step-2.

10<sup>7</sup> cells. Avoid excessive amount)

ars, and centrifuged at 300 × g in a tube for 5 minutes to form nt completely,, proceed to Step-2.

ompany) of 700μL, and suspend the cells by vortex mixer. emperature (20-30 °C ).

ap the tube gently to loosen the particles before adding Lysis

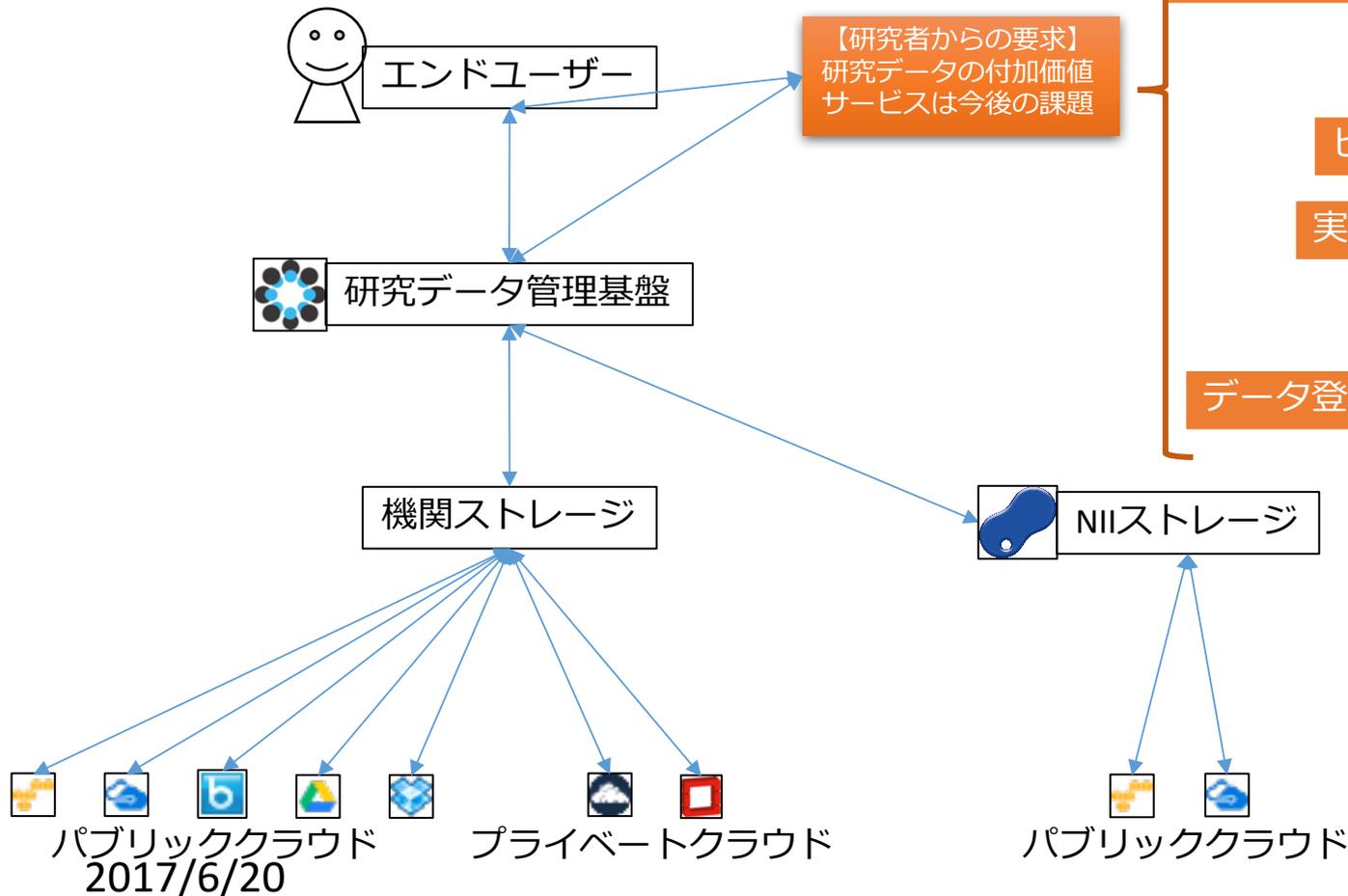
leep in  
ains to

## ファイル管理と閲覧

- ファイルのバージョン管理
- 多様なファイル形式に対応したプレビュー

2017/6/20

## 恒常的な安定運用の実現と 研究データ処理の高度化



データ管理計画の登録

データ公開の承認プロセスの多重化

情報セキュリティポリシーへの対応

データ解析環境の共有

ビッグデータ/HPC対応

実験装置データの自動集約

研究データの教育利用

データ登録インセンティブの向上

# 結言

---

- 米国ではCOSのOSFはロングテールデータを中心にセントラルサービスとしてデータ基盤の利用や引用が拡大しています。一方で先進的な大学では自身でそれを構築しています。
- ヨーロッパではEUDAT CDIの機関を中心に、各国各分野の研究機関を結ぶe-scienceのインフラの上にデータ基盤サービスEUDATが構成されています。
- 国立情報学研究所では研究データの**検索・公開・管理**の3種類の**データ基盤**サービスを、大学・研究向けに開発しています。そのうち**管理基盤**は既存の機関が持つクラウドストレージや研究ツールのハブとして利用いただけるWebアプリケーションです。