

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

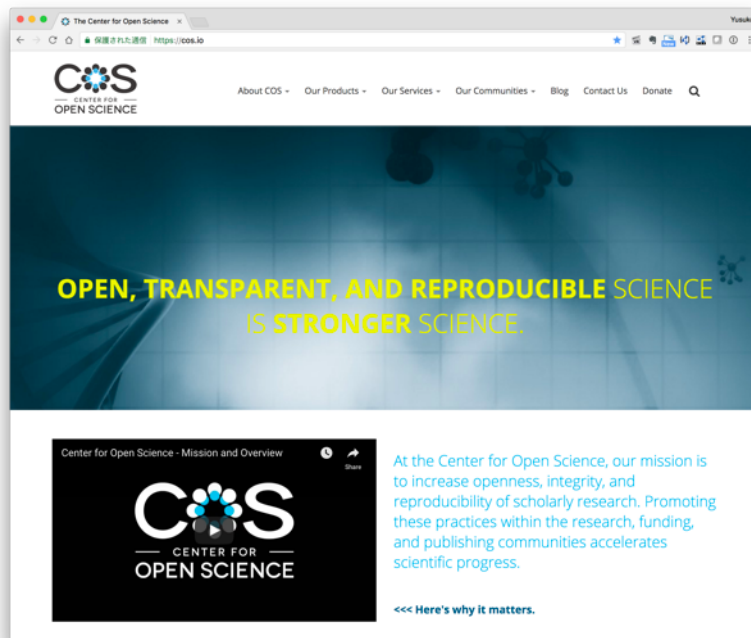
# 日本と欧米の研究データ基盤

---

- 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

# 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.

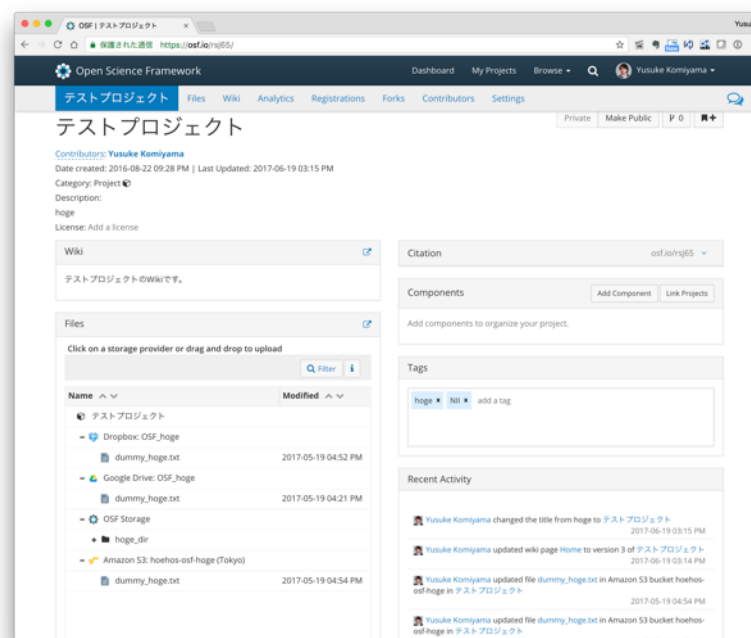
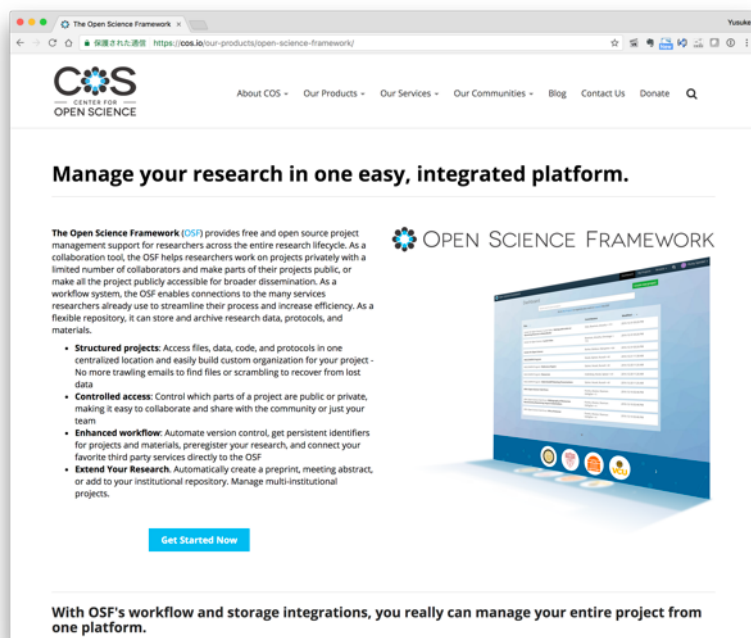
2017/6/20

# 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.

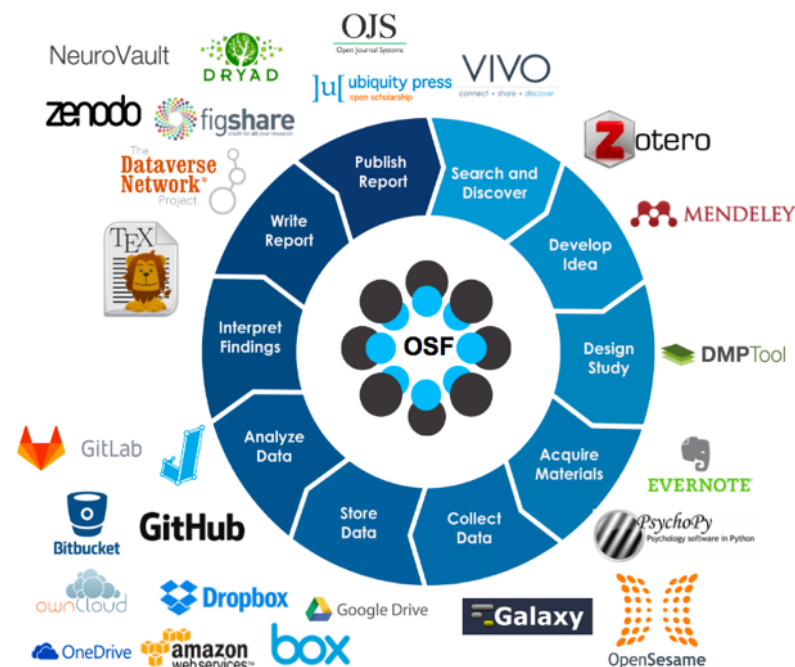
<https://cos.io/our-products/open-science-framework/>

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



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

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

2017/6/20

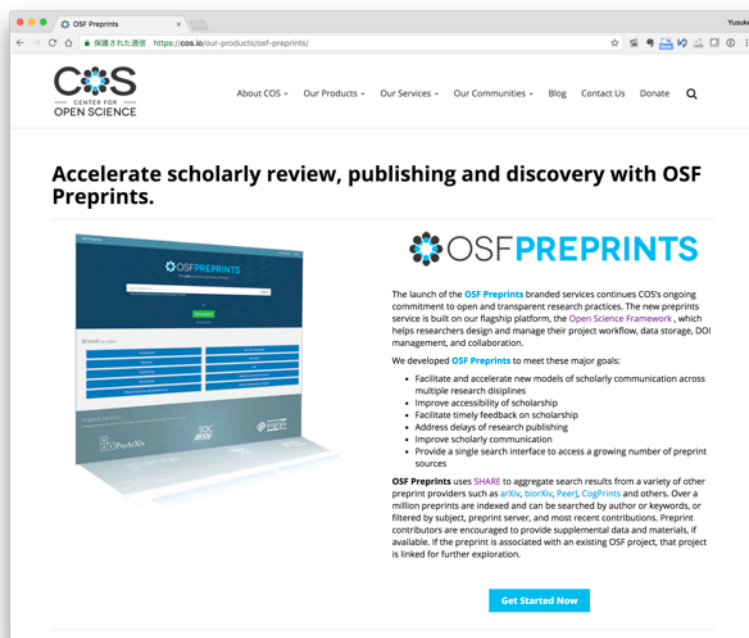
# OSFを用いたプレプリントの展開

## 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



### 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/>

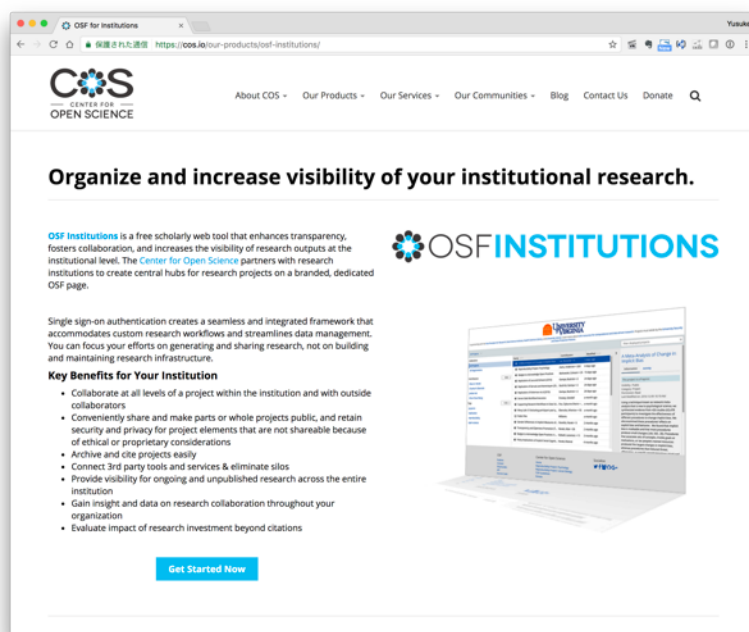
2017/6/20

# cosのその他のサービス

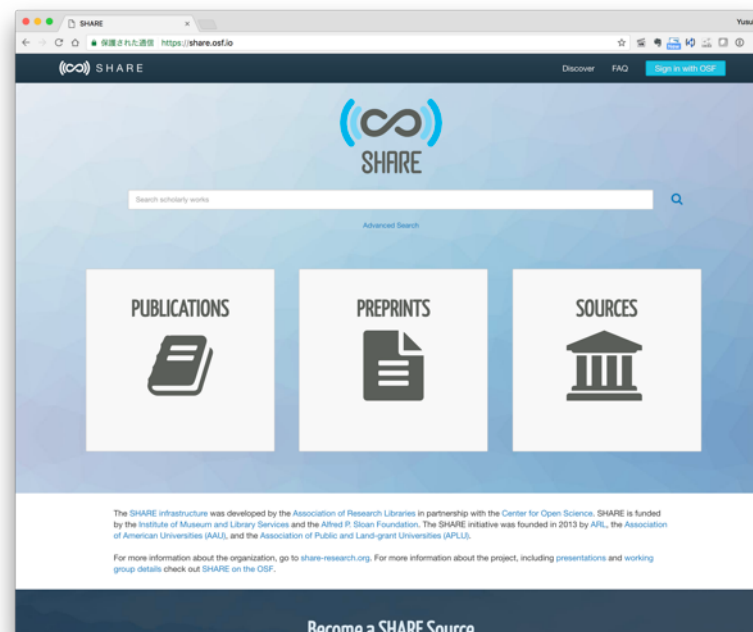
## OSF INSTITUTIONS

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

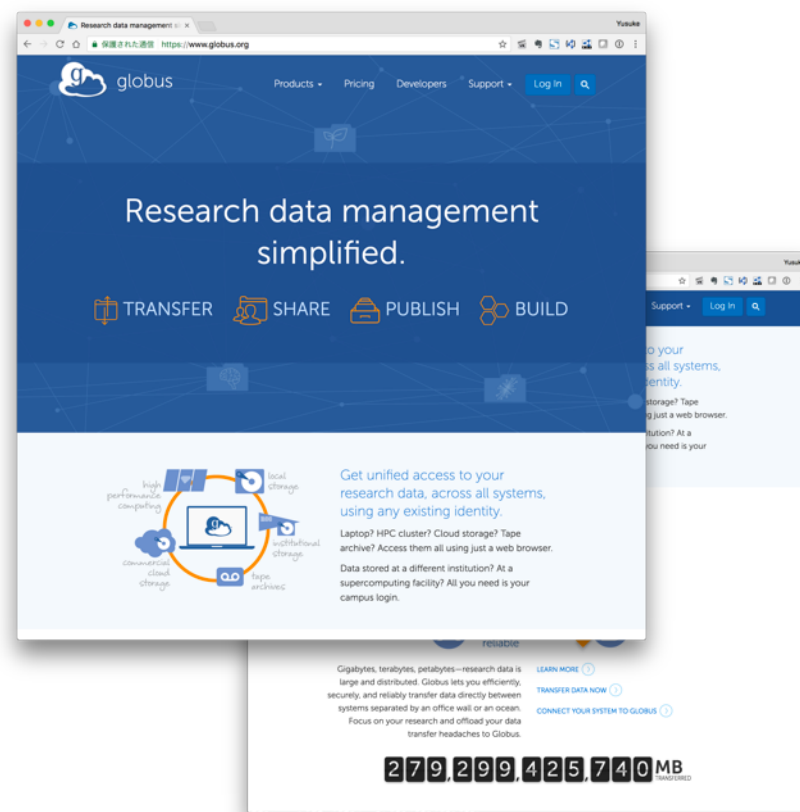
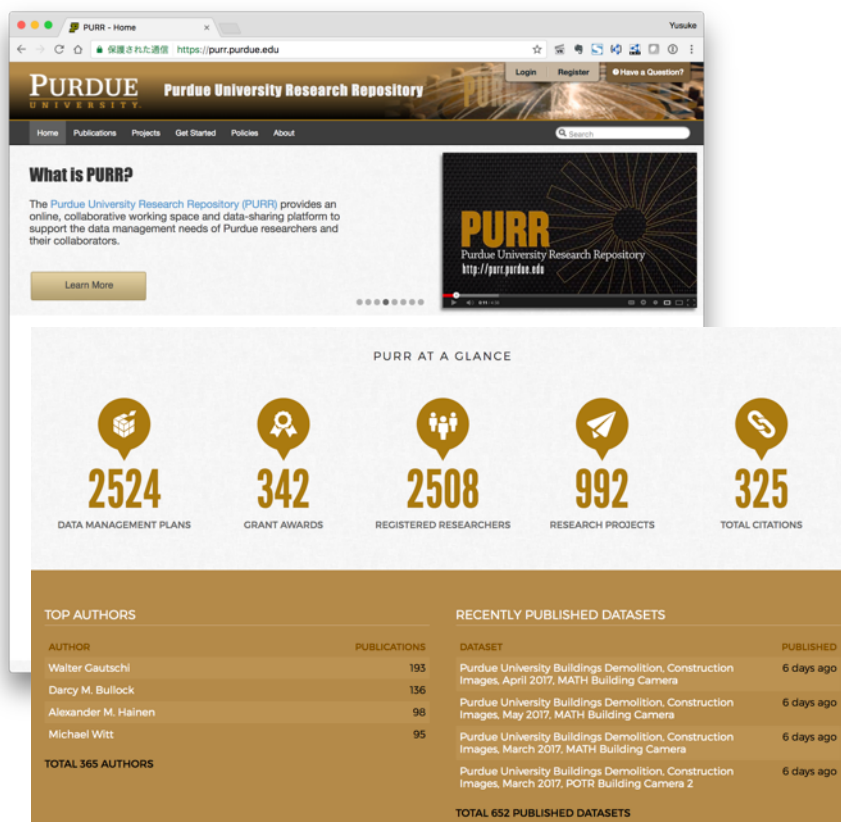
PURR: Purdue University Research Repository

図書館系

globus

e-Science系

(The University of Chicago and Argonne National Laboratory)



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

2017/6/20

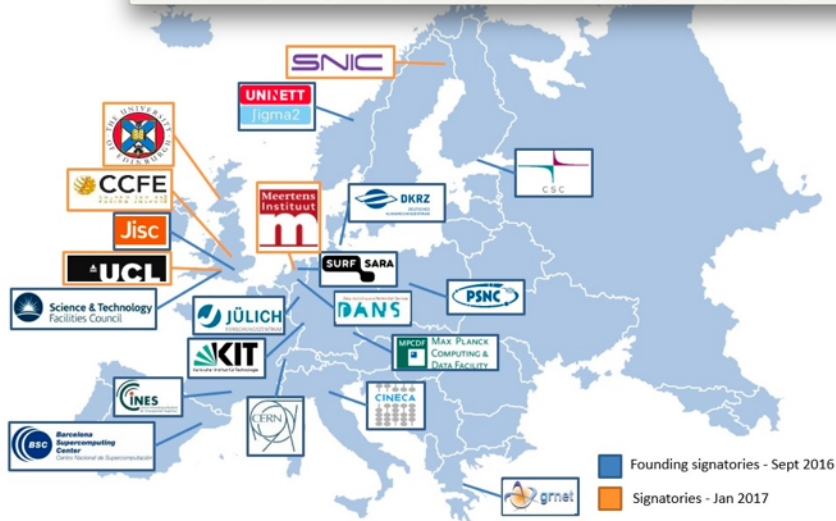
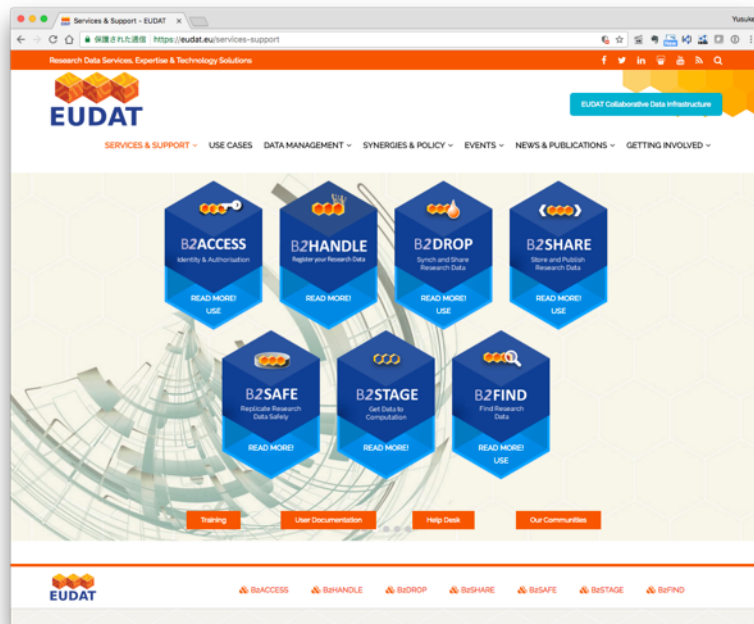
<https://purrr.purdue.edu/>

R. Ananthakrishnan, 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.

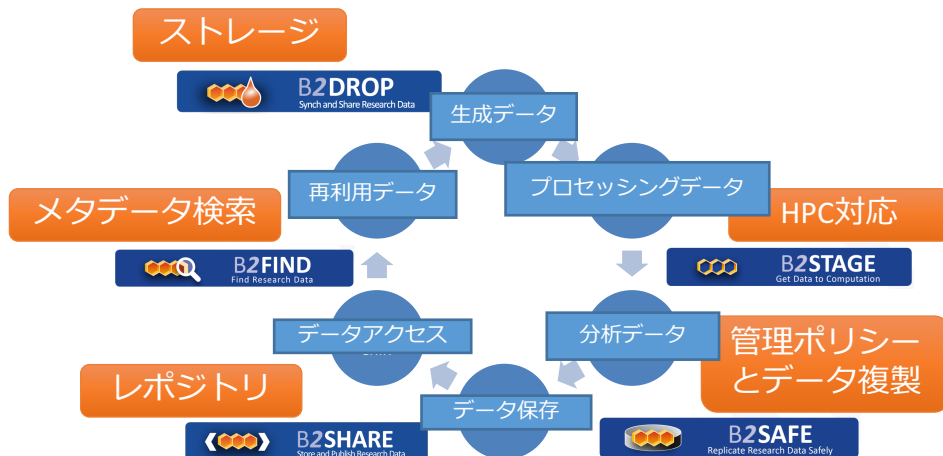
A. S. Memon, et al., in 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing, 2014, pp. 726–731.

# EUDATサービス

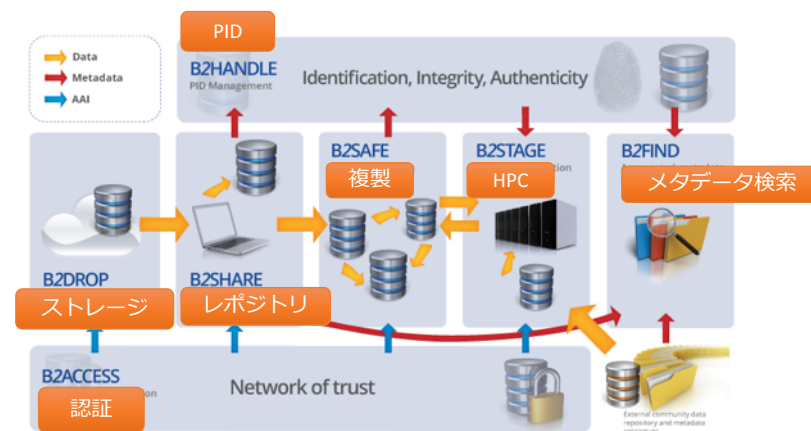
研究におけるデータ利用のサイクルと  
初期のEUDATAサービス



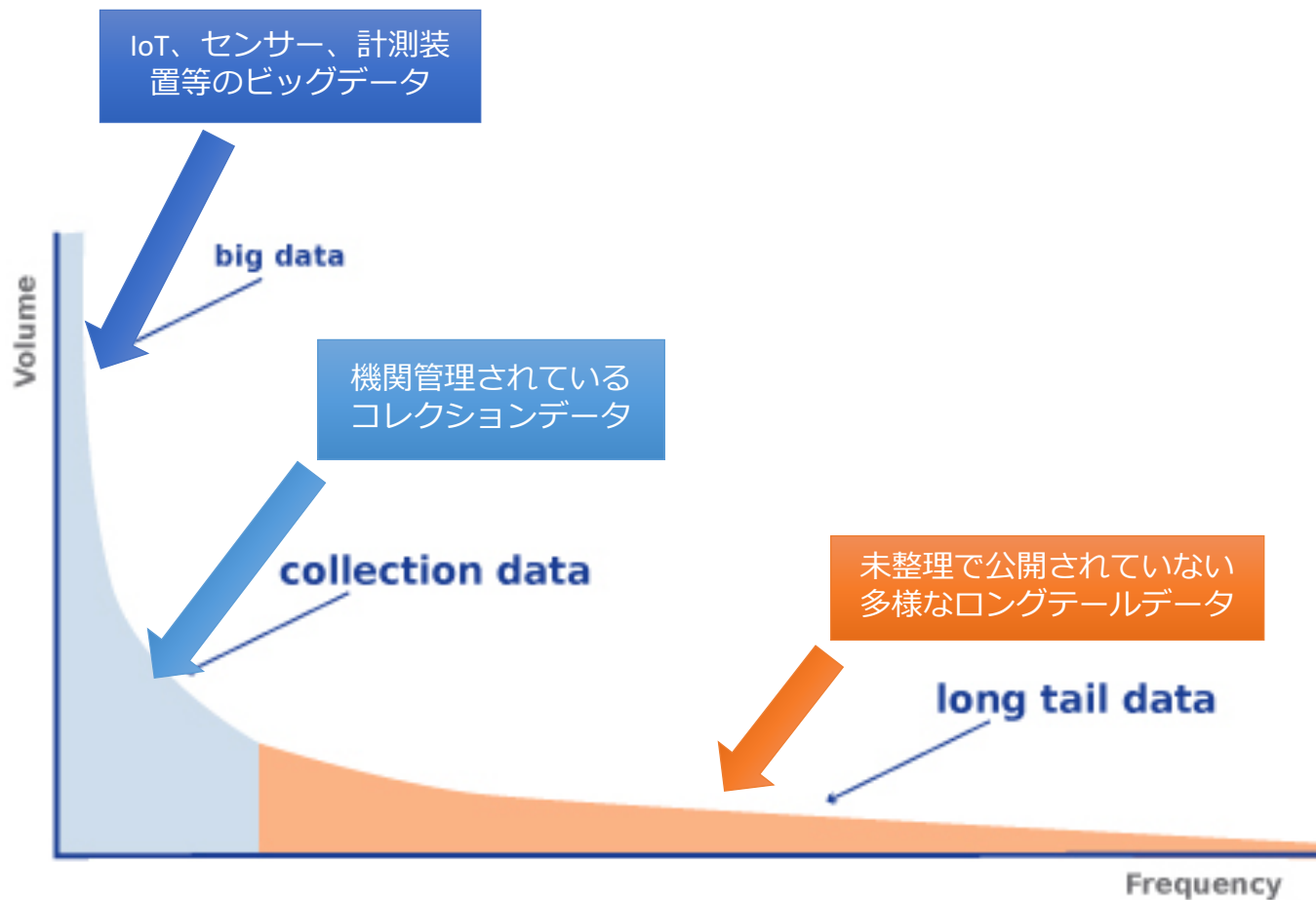
Support throughout the lifecycle



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>

# 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



学術認証  
フェデレーション

HPCI認証



無線LAN  
ローミング

### 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



電子証明書



VPN

### 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



超高速・高機能回線

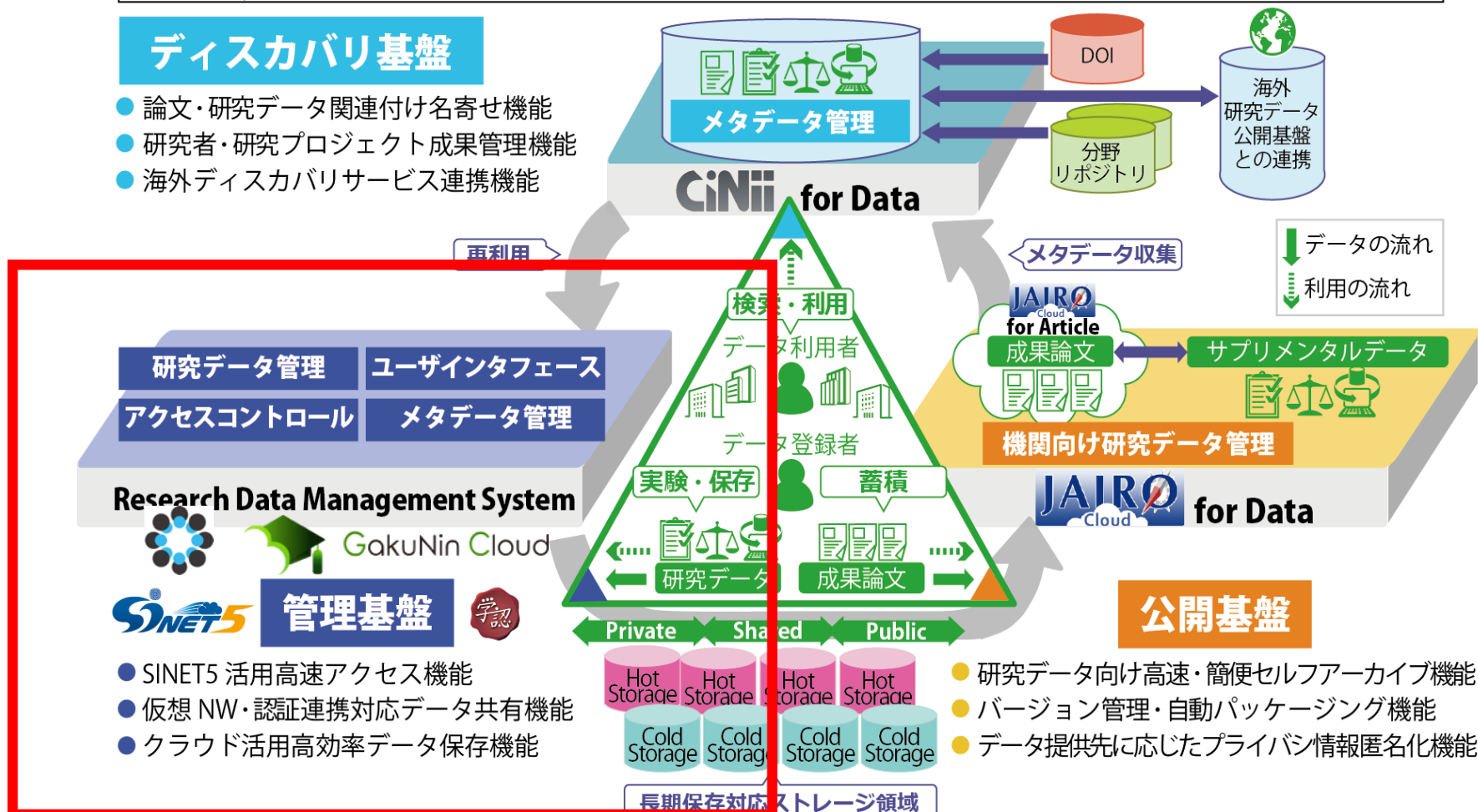
アクセス回線共同調達

# 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



「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

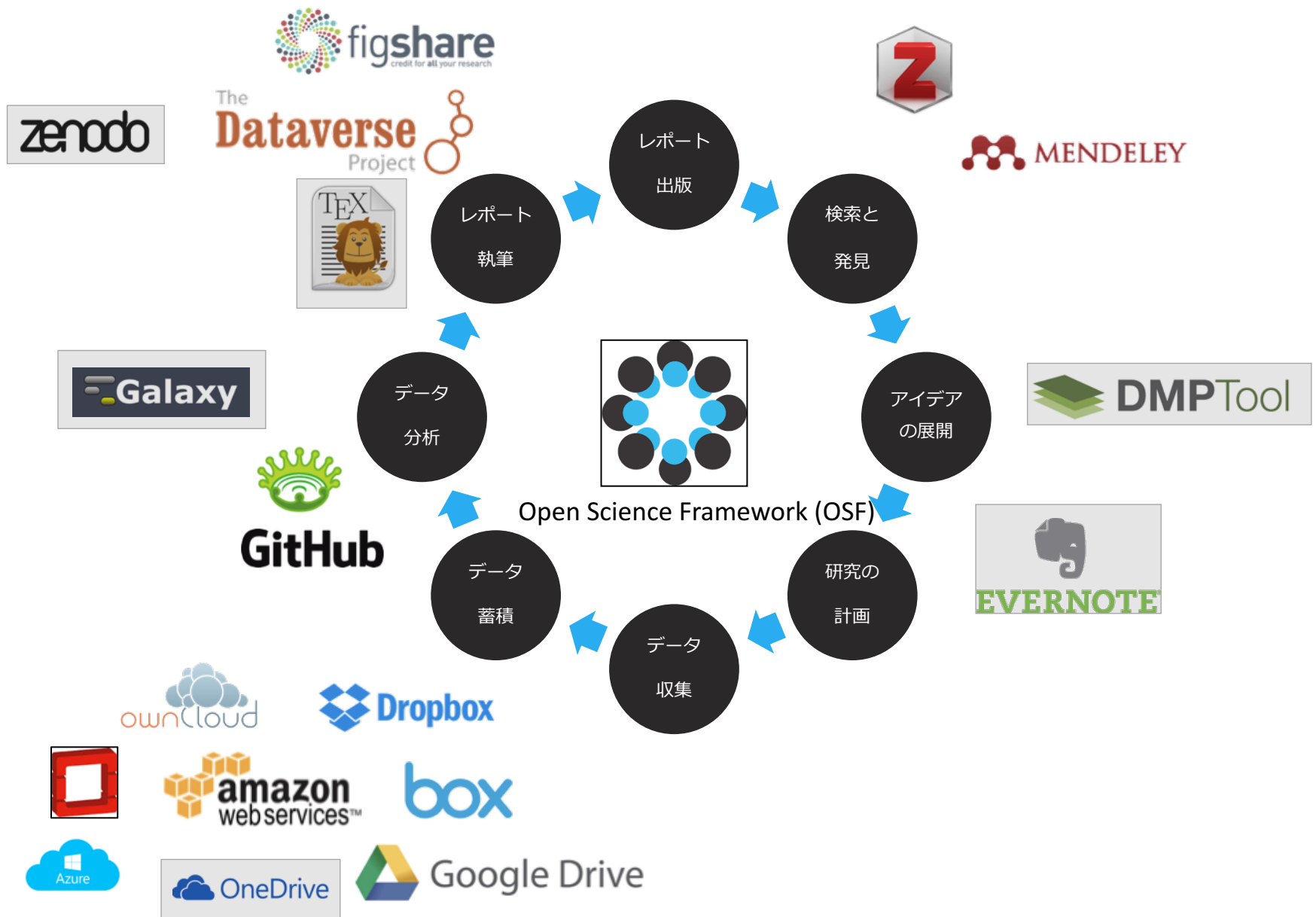


# 研究目的

- 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。

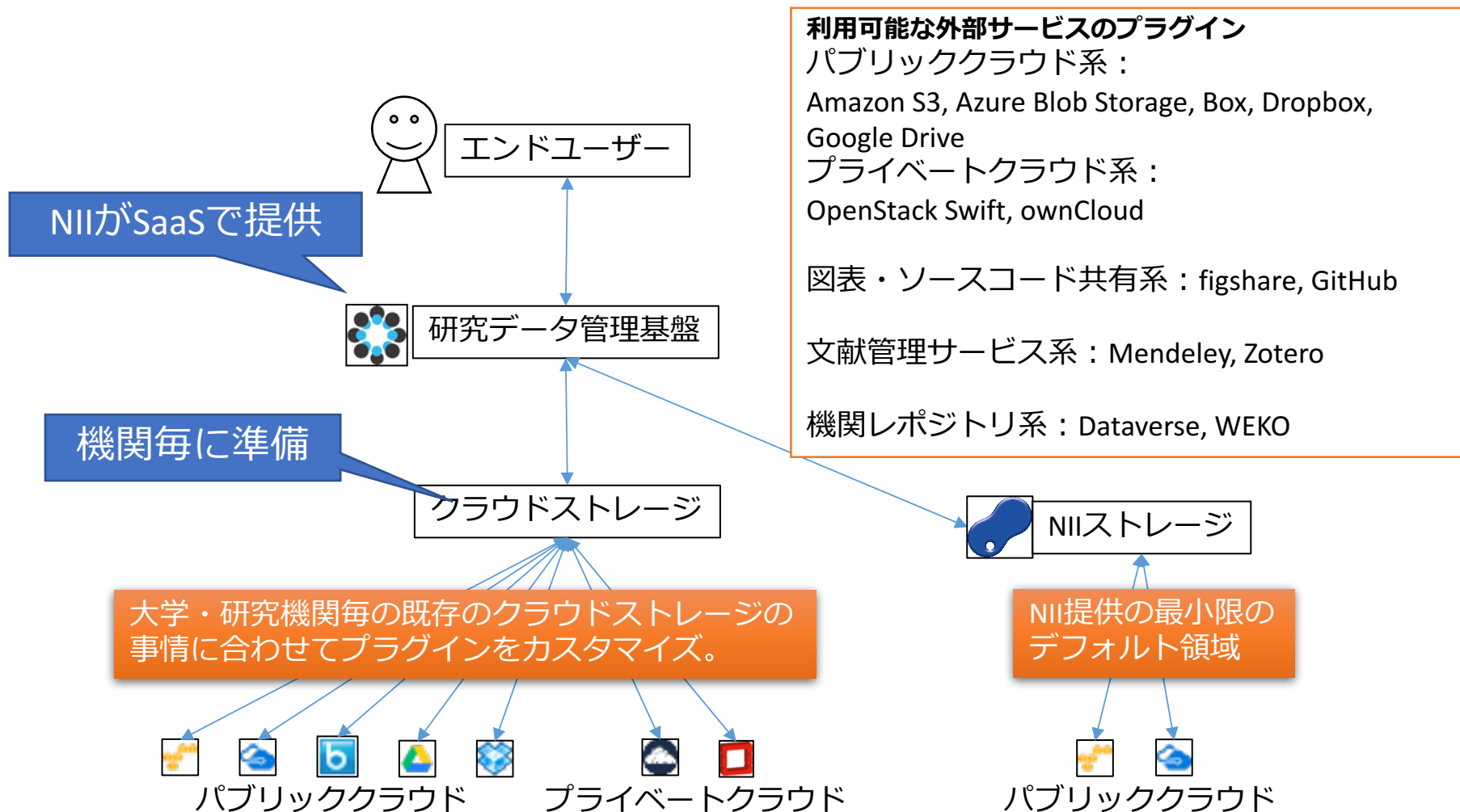
# 研究者が日常的に使う研究ツールのハブ

国立情報学研究所  
National Institute of Informatics

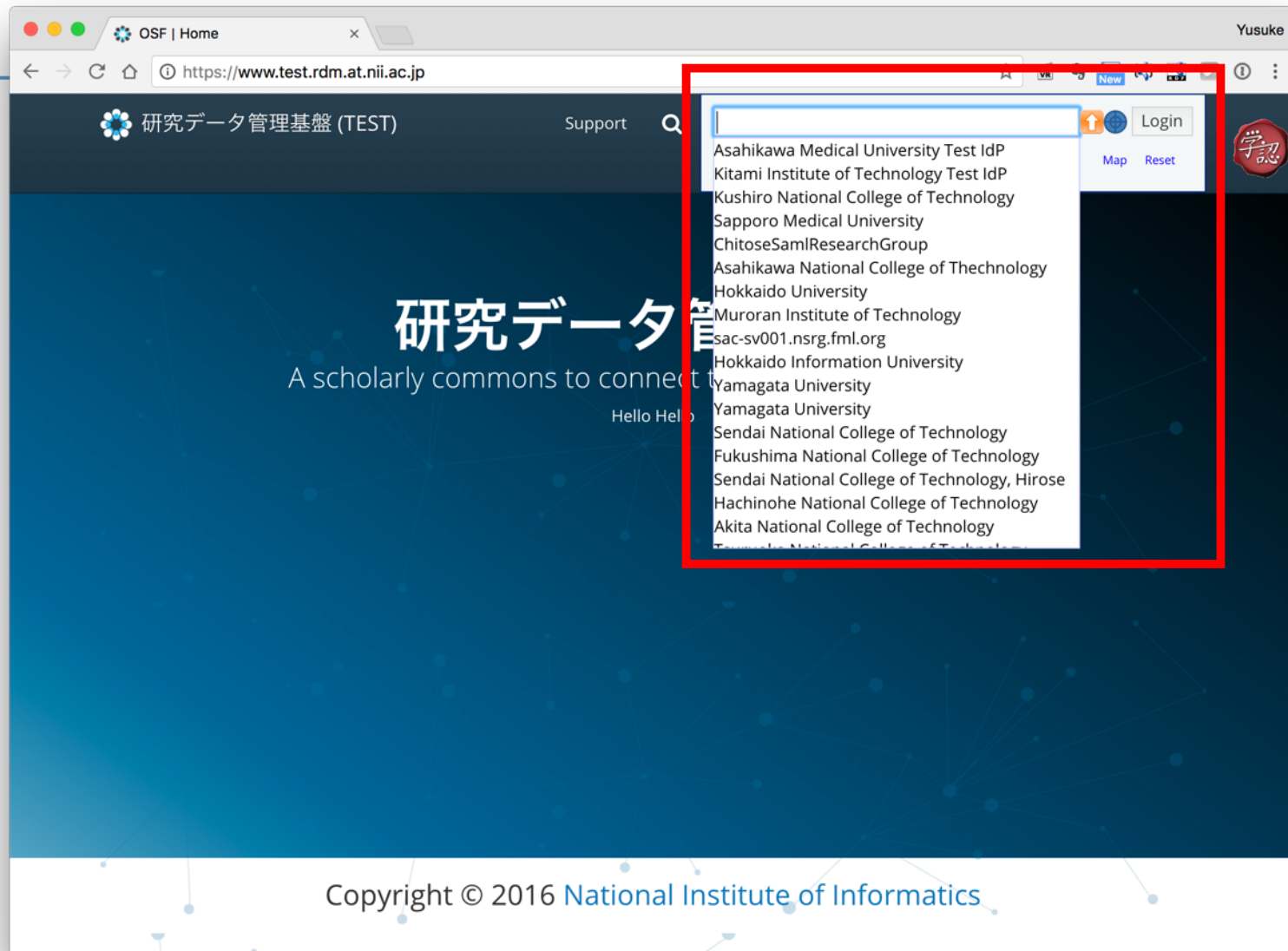


“Opening Science by Opening Workflow”, by Jeff Spies: <https://cos.io/pr/2015-09-24/> より

## 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



GakuNin DS

Open Science Frameworkを学認へ対応



# 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

Collections

- All my projects
- All my registrations
- Bookmarks (0)

Contributors

- Satoshi Yazawa

Tags

- nii
- osf
- hoge
- 国立情報学研究所

Name	Contributors	Modified
hoge	Komiyama	4 days ago
OSF	Komiyama	11 days ago
yazawa-test	Yazawa, Komiyama	16 days ago
hoge hoge2	Komiyama	16 days ago
Fork of Fork of hoge	Komiyama	17 days ago
Fork of hoge	Komiyama	17 days ago

Filter displayed projects

hoge

Information Activity

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

hoge

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

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 Q 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 Open Science Framework (OSF) interface. At the top, there's a navigation bar with 'Open Science Framework' and user information 'Yusuke Komiyama'. Below this, a secondary navigation bar shows tabs for 'hoge', 'Files', 'Wiki', 'Analytics', 'Registrations', 'Forks', 'Contributors', and 'Settings'. The main content area is titled 'Click on a storage provider or drag and drop to upload'. It features a toolbar with buttons for 'Upload', 'Create Folder', 'Delete Folder', 'Download as zip', 'Rename', 'Filter', and 'Info'. A table lists files and folders under the 'hoge' project, organized by storage provider (OSF Storage and Amazon S3). The table columns are Name, Size, Version, Download..., and Modified. A red callout box points to the 'dummy\_hoge.txt' file in the Amazon S3 section.

Name	Size	Version	Download...	Modified
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

Filter ^

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  
roceed 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

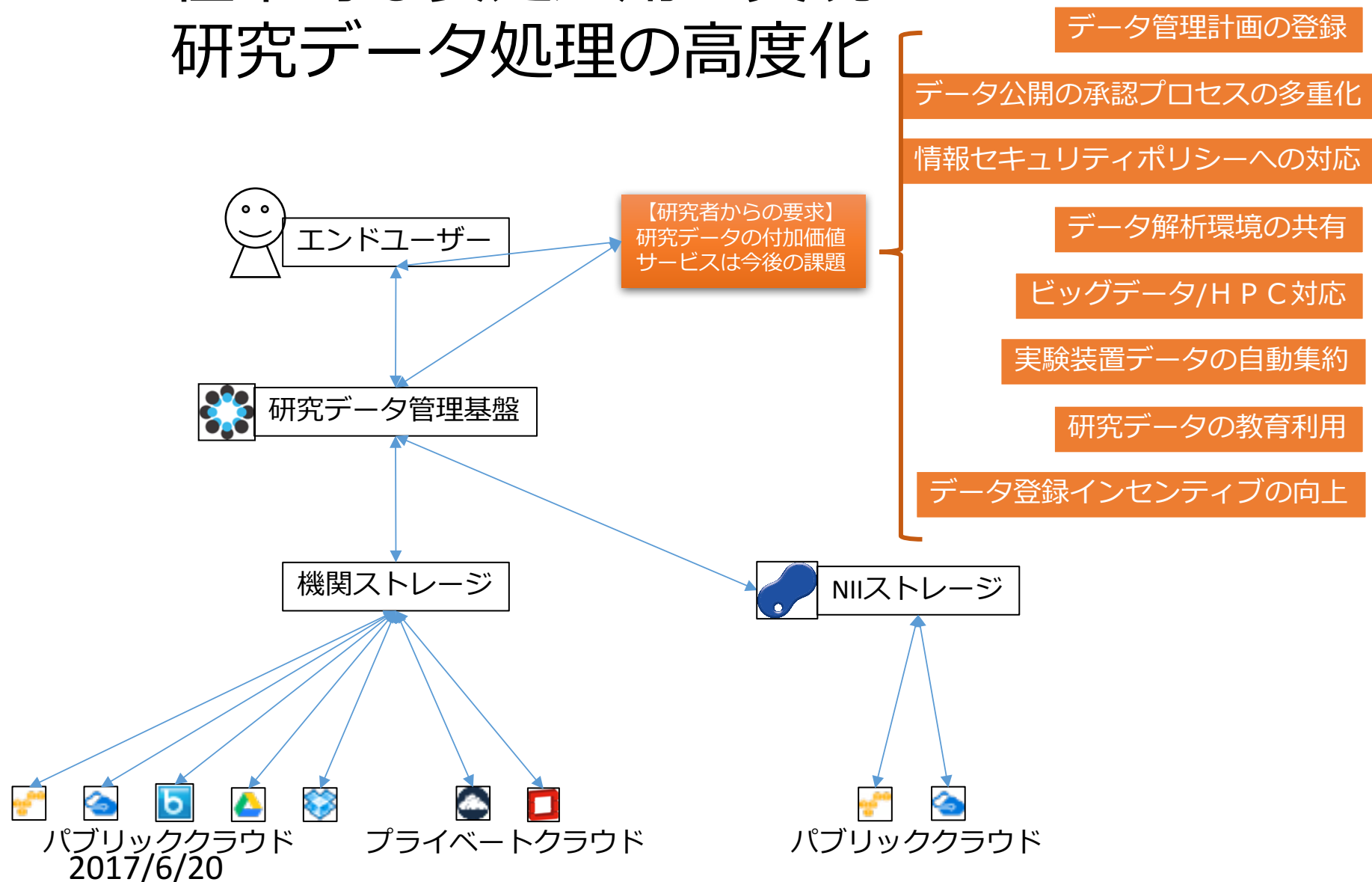
leep in  
ains to

## 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



# 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。