

オンライン分析システムの 開発状況と活用事例

藤原一毅 林正治 朝岡誠 船守美穂 山地一禎
国立情報学研究所 オープンサイエンス基盤研究センター

2021/09/07 統計関連学会連合大会
人文学・社会科学分野におけるデータインフラストラクチャー構築推進事業の展開

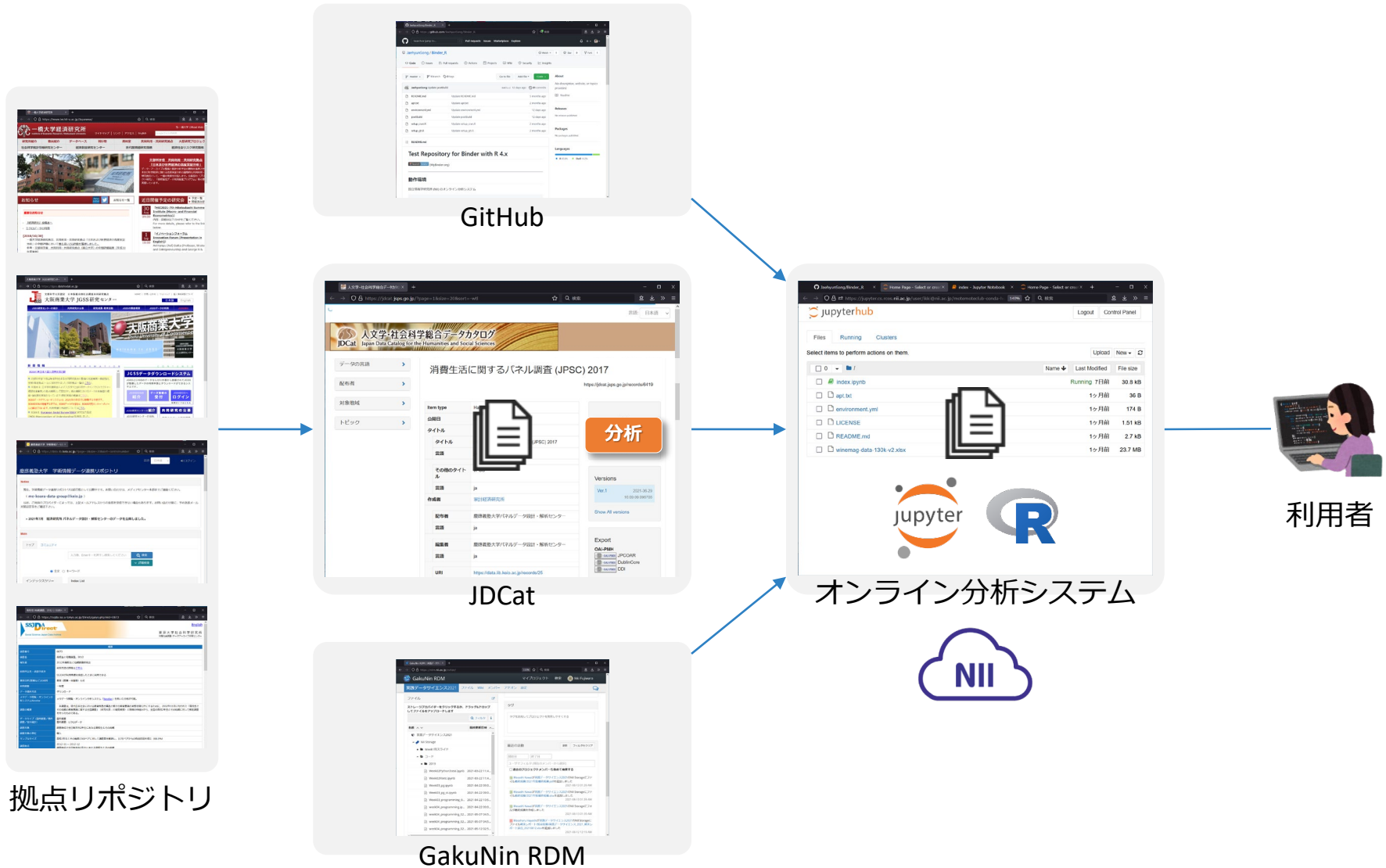
まとめ

- 国立情報学研究所 (NII) は、JSPS人文学・社会科学データインフラストラクチャー構築推進事業の一環として「**オンライン分析システム**」を開発しています。
- **R**や**Python**で任意のプログラムが書ける分析環境を提供します。
 - 現在、Jupyter Notebook, JupyterLab, RStudio が利用可能です。
- 学術認証フェデレーション「**学認**」の参加機関に所属する教職員や学生が利用できます。
- JDCatで公開されているデータを1クリックで取り込み、すぐに分析できるようになります。（開発中）
- 二次分析を始めるハードルが下がり、分野横断的な研究の促進につながります。

目次

- オンライン分析システムとは
 - システム構成
 - ユースケース
- 単体利用／GitHubとの連携
 - 授業での先行試用
 - デモ
 - 協力者募集中
 - 類似サービスとの比較
- JDCatとの連携（開発中）
 - 制限公開データの保護
- GakuNin RDMとの連携（開発中）
- まとめ

システム構成



ユースケース

研究

- ご自身の研究のためのデータ分析



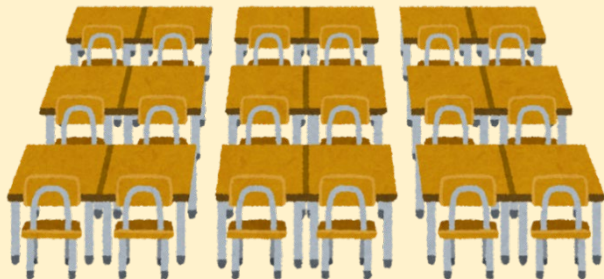
公開・共有

- 他の研究者の二次分析に資するデータとプログラムの公開



教育・学習

- 学生たちにデータ分析をさせるゼミ・講義・演習など



引き継ぎ

- 先輩の研究環境を後輩が再現し、研究を継続する



単体利用／GitHubとの連携

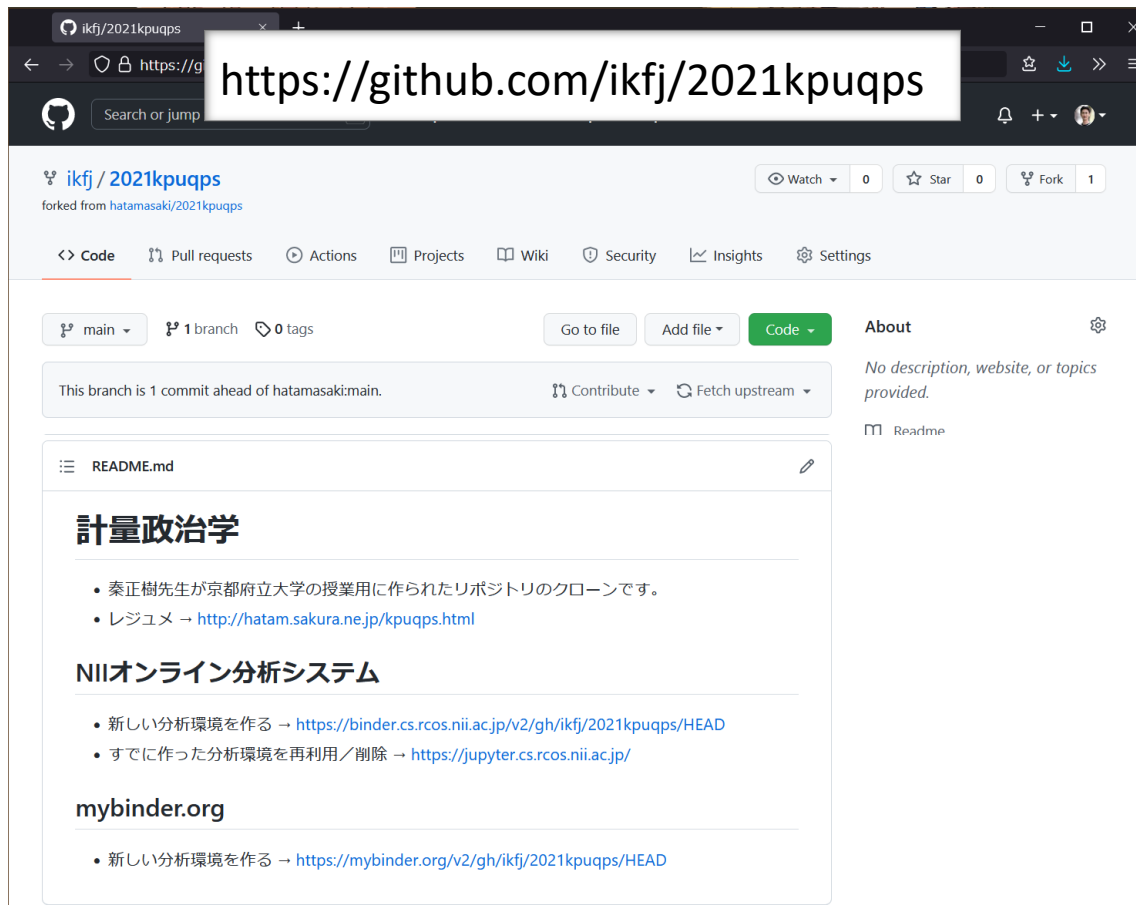


授業での先行試用

講師	機関	IdP	科目	人数	時期	リポジトリ
秦先生	京都府立大	kpu.ac.jp	政治学3 (計量政治学)	6	前期	https://github.com/hatamasaki/2021kpuqps
	関西大	kansai-u.ac.jp	展開講義 (政治学方法論)	4	前期	
宋先生	関西大	kansai-u.ac.jp	ミクロ政治データ分析実習	44	前期	https://github.com/JaehyunSong/Binder_R
			マクロ政治データ分析実習	?	後期	
	京都府立大	kpu.ac.jp	情報処理応用実習	10	集中	
	神戸大	OpenIdP	計量政治学方法論I (理論と実証)	?	集中	
浅野先生	早稲田大	OpenIdP	計量分析 (政治01)	61	前期	https://github.com/ikfj/ma-keiryō
			計量分析 (政治02)	60	後期	
			計量分析 (政治03)	60	後期	
			政治分析A	27	前期	
			社会調査法	2	前期	
			Methods of Social Survey	20	後期	
	拓殖大	OpenIdP	(ゼミ)	25	前期	https://github.com/ikfj/ma-keiryō
林田先生	北九州市立大	OpenIdP	(ゼミ)	11	後期	https://github.com/motemoteclub/conda
永崎先生	東京大	u-tokyo.ac.jp	(講義)	?	後期	GakuNin RDM
	広島大	hiroshima-u.ac.jp	(集中講義)	?	後期	GakuNin RDM

デモ

- 関西大、立教大、京都府立大、広島大、東京大の方は、大学のIDでNIIのオンライン分析システムをご試用いただけます。
- その他の方は、mybinder.org で使用感をお試しください。



https://github.com/ikfj/2021kpuqps

ikfj / 2021kpuqps
forked from hatamasaki/2021kpuqps

Code Pull requests Actions Projects Wiki Security Insights Settings

main 1 branch 0 tags
Go to file Add file Code

This branch is 1 commit ahead of hatamasaki:main. Contribute Fetch upstream

README.md

計量政治学

- 秦正樹先生が京都府立大学の授業用に作られたリポジトリのクローンです。
- レジューメ → <http://hatam.sakura.ne.jp/kpuqps.html>

NIIオンライン分析システム

- 新しい分析環境を作る → <https://binder.cs.rcos.nii.ac.jp/v2/gh/ikfj/2021kpuqps/HEAD>
- すでに作った分析環境を再利用/削除 → <https://jupyter.cs.rcos.nii.ac.jp/>

mybinder.org

- 新しい分析環境を作る → <https://mybinder.org/v2/gh/ikfj/2021kpuqps/HEAD>



協力者募集中

- 先行試用にご協力いただける方を募集しています！

NII オンライン分析システム

検索



スクリーンショット: NII オンライン分析システム (実証実験) の Confluence ページ

ブラウザ: https://meatwiki.nii.ac.jp/confluence/pages/viewpage.action?pageId=48137275

検索: [検索] ログイン

ページ: **オンライン分析システム (実証実験)**
confluence-admin が作成し、2021/05/25 に ikki が最終更新

English / 日本語

受講者ガイド
本システムを利用して行われる授業やセミナー等を受講する方は、[こちら](#)をお読みください。

講師ガイド
本システムを利用する研究者や教員・講師の方は、[こちら](#)をお読みください。

OpenIDP の使い方
OpenIDP を初めて利用する方は、[こちら](#)をお読みください。

情報基盤担当者ガイド
本システムを利用する機関で「学認」および学外ネットワークを担当する部署の方は、[こちら](#)をお読みください。

オンライン分析システムとは

- Jupyter Notebook と RStudio の実行環境を NII がクラウド上で提供するものです。
- 研究者は、オンライン分析システムにログインして、すぐに R や Python のプログラムを作成・実行することができます。
- 共同研究者や学生は、公開されている研究データとプログラムをボタンひとつでオンライン分析システムに取り込み、すぐにデータ分析を始めることができます。
- ゼミ・講義・演習など、多くの受講者に分析環境を提供する用途にも適しています。環境構築にかかる教員・講師の皆さまの負担が軽減されます。

実証実験の概要

- [こちらのスライド](#)をご参照ください。

参加要件

- 学術認証フェデレーション「学認」に参加している機関（大学等）に

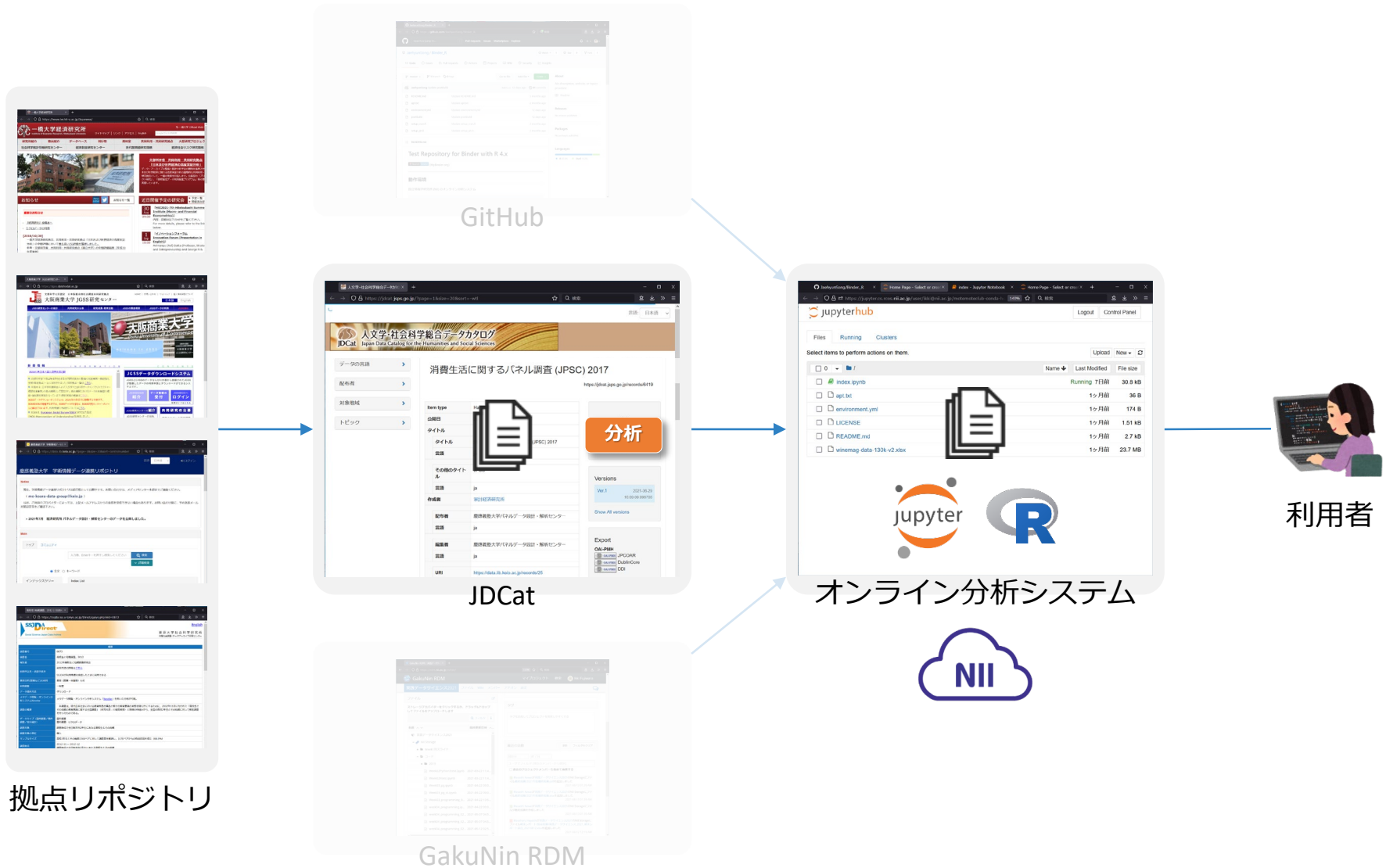
国立情報学研究所 (NII) は、日本学術振興会 (JSPS) 人文・社会科学データインフラストラクチャー構築推進事業の一環として「オンライン分析システム」を開発しています。

本システムは2020年10月から実証実験（クローズドβテスト）を開始し、現在、実験にご参加いただける大学・研究機関等を募集しています。ご興味のある方は下記担当者までお問い合わせください。

類似サービスとの比較

	本サービス	GESIS Notebooks	mybinder.org	Google Colab	Pangeo's BinderHub	Codalab Worksheets
対象分野	汎用	社会科学	汎用	主に深層学習	地球科学	深層学習
提供元	NII (日・学術機関)	GESIS (独・学術機関)	Project Jupyter (任意団体)	Google (米・民間企業)	Pangeo (米・学術団体)	Microsoft (米・民間企業)
アカウント	学認	GESIS	不要	Google	GitHub	Codalab
対応言語	Python, R, Julia	Python, R, Julia	Python, R, Julia	Python, R, Julia, Swift	Python	Python
対応リポジトリ	WEKO3 (JDCat), GakuNin RDM, GitHub, Gist, GitLab, Zenodo, Figshare, Hydroshare, Dataverse	GitHub, Gist, GitLab, Zenodo, Figshare, Hydroshare, Dataverse	GitHub, Gist, GitLab, Zenodo, Figshare, Hydroshare, Dataverse	Google Drive, GitHub	GitHub, Gist, GitLab, Zenodo	?
メモリ ストレージ	3GB 10GB	32GB 10GB	2GB ?	13GB 40GB	8GB ?	?
タイムアウト	30日 / 永続	40分	10分	90分 / 12時間	11分 / 3時間	?
インフラ	オンプレ	オンプレ	Google, OVH, Turing Institute	Google	Google	Microsoft

JDCatとの連携



JDCatとの連携（開発中）

The screenshot shows a web browser window displaying the JDCat website. The address bar shows the URL <https://jdcatalog.jsps.go.jp/records/7533>. The page header includes the JDCat logo and the text '人文学・社会科学総合データカタログ Japan Data Catalog for the Humanities and Social Sciences'. On the left side, there are navigation buttons for 'データの言語', '配布者', '対象地域', and 'トピック'. The main content area displays the title '馬車軌道 (内地) (昭和1-10年) : 大日本帝国統計年鑑 56 (昭和12年) 表168' and the URL <https://jdcatalog.jsps.go.jp/records/7533>. Below the title is a table with metadata:

Item type	Harvesting DDI(1)
公開日	2021-07-29
タイトル	
タイトル	馬車軌道 (内地) (昭和1-10年) : 大日本帝国統計年鑑 56 (昭和12年) 表168
言語	ja
作成者	鉄道省
配布者	一橋大学経済研究所
言語	ja
配布者URI	https://www.ier.hit-u.ac.jp/Japanese/
編集者	内閣統計局
言語	ja
URI	https://doi.org/10.50914/0002001612
URI	https://d-repo.ier.hit-u.ac.jp/records/2001612

On the right side, there is a '3 views' indicator and an orange '分析' (Analysis) button. Below that is a 'Versions' section showing 'Ver.1' with a timestamp of '2021-07-09 18:28:20.796921' and a 'Show All versions' link. At the bottom right, there is an 'Export' section with options for OAI-PMH, JPCOAR, DublinCore, and DDI.

JDCatとの連携 (開発中)

Home Page - Select or creat × Untitled - Jupyter Notebook × RStudio × +

← → https://jupyter.cs.rcos.nii.ac.jp/user/jikki@nii.ac.jp/ikfj-ds-polimetrics-19gj4vvp/n ☆ 🔍 検索 🏠 ⬇️ ⋮

jupyterhub Untitled Last Checkpoint: 5分前 (unsaved changes) Logout Control Panel

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 O

Run Code Download Memory: 238 / 3072 MB

GitHub Binder

```
In [2]: class PDF(object):
def __init__(self, pdf, size=(200,200)):
self.pdf = pdf
self.size = size

def _repr_html_(self):
return '<iframe src={0} width={1[0]} height={1[1]}></iframe>'.format(self.pdf, self.size)

def _repr_latex_(self):
return r'\includegraphics[width=1.0\textwidth]{{{0}}}'.format(self.pdf)
```

In [5]: PDF('1937_2_07_168.pdf', size=(800,600))

Out[5]:

296 297

163. 鉄道営業収支 (単位:千円)

年次	国鉄		私鉄		計
	営業収入	営業支出	営業収入	営業支出	
昭和1	484,030	275,376	213,246	7,915	697,271
2	324,540	201,960	224,800	11,914	546,454
3	324,102	200,706	229,414	7,301	561,523
4	317,070	204,442	212,632	4,491	538,635
5	493,140	304,234	173,391	4,441	676,306
6	435,641	304,038	164,361	4,167	604,147
7	423,094	284,032	160,373	4,449	588,948
8	472,564	301,331	172,264	5,241	651,360
9	513,646	314,126	214,543	5,461	728,776
10	574,996	359,637	218,496	6,471	799,590
計	99,470	45,989	34,982	5,619	124,829
増減	2,872	14,653	12,954	58	28,547

164. 地方鉄道概況 (内訳)

年次	線区数	営業収入		営業支出		営業収入	営業支出
		営業収入	営業支出	営業収入	営業支出		
昭和1	1	2,237,000	2,270	28,512	...	2,237,000	2,270
2	2	56,209	2,311	2,600	...	56,209	2,311
3	3	31,741	2,406	2,396	...	31,741	2,406
4	4	43,387	2,697	3,023	...	43,387	2,697
5	5	42,865	3,457	3,363	...	42,865	3,457
6	6	42,203	2,410	2,491	...	42,203	2,410
7	7	44,821	2,362	2,472	...	44,821	2,362
8	8	41,352	2,272	2,171	...	41,352	2,272
9	9	42,846	2,324	2,343	...	42,846	2,324
10	10	45,621	2,135	2,195	...	45,621	2,135
計
増減

165. 鉄道事故

年次	線区数	人身		貨物	
		人身	貨物	人身	貨物
昭和1	1	63	1,234	1,001	1,291
2	2	63	1,234	1,001	1,291
3	3	63	1,234	1,001	1,291
4	4	63	1,234	1,001	1,291
5	5	63	1,234	1,001	1,291
6	6	63	1,234	1,001	1,291
7	7	63	1,234	1,001	1,291
8	8	63	1,234	1,001	1,291
9	9	63	1,234	1,001	1,291
10	10	63	1,234	1,001	1,291
計
増減

166. 電気軌道 (内訳)

年次	会社数	乗客		乗車		乗客	乗車
		乗客	乗車	乗客	乗車		
昭和1	1	1,839	1,023	4,222	778	1,272,272,670	114,714,444
2	2	99	1,064	1,090	4,373	723	1,207,116,288
3	3	99	1,064	1,090	4,373	723	1,207,116,288
4	4	101	1,065	1,091	4,374	724	1,208,117,291
5	5	101	1,065	1,091	4,374	724	1,208,117,291
6	6	101	1,065	1,091	4,374	724	1,208,117,291
7	7	101	1,065	1,091	4,374	724	1,208,117,291
8	8	101	1,065	1,091	4,374	724	1,208,117,291
9	9	101	1,065	1,091	4,374	724	1,208,117,291
10	10	101	1,065	1,091	4,374	724	1,208,117,291
計
増減

167. 汽船軌道 (内訳)

年次	会社数	乗客		乗車		乗客	乗車
		乗客	乗車	乗客	乗車		
昭和1	1	593	418	329	663	11,222,042	1,202,288
2	2	437	424	339	662	11,222,042	1,202,288
3	3	437	424	339	662	11,222,042	1,202,288
4	4	437	424	339	662	11,222,042	1,202,288
5	5	437	424	339	662	11,222,042	1,202,288
6	6	437	424	339	662	11,222,042	1,202,288
7	7	437	424	339	662	11,222,042	1,202,288
8	8	437	424	339	662	11,222,042	1,202,288
9	9	437	424	339	662	11,222,042	1,202,288
10	10	437	424	339	662	11,222,042	1,202,288
計
増減

168. 馬車軌道 (内訳)

年次	会社数	乗客		乗車		乗客	乗車
		乗客	乗車	乗客	乗車		
昭和1	1	166	181	107	888	316	1,066,264
2	2	166	181	107	888	316	1,066,264
3	3	166	181	107	888	316	1,066,264
4	4	166	181	107	888	316	1,066,264
5	5	166	181	107	888	316	1,066,264
6	6	166	181	107	888	316	1,066,264
7	7	166	181	107	888	316	1,066,264
8	8	166	181	107	888	316	1,066,264
9	9	166	181	107	888	316	1,066,264
10	10	166	181	107	888	316	1,066,264
計
増減

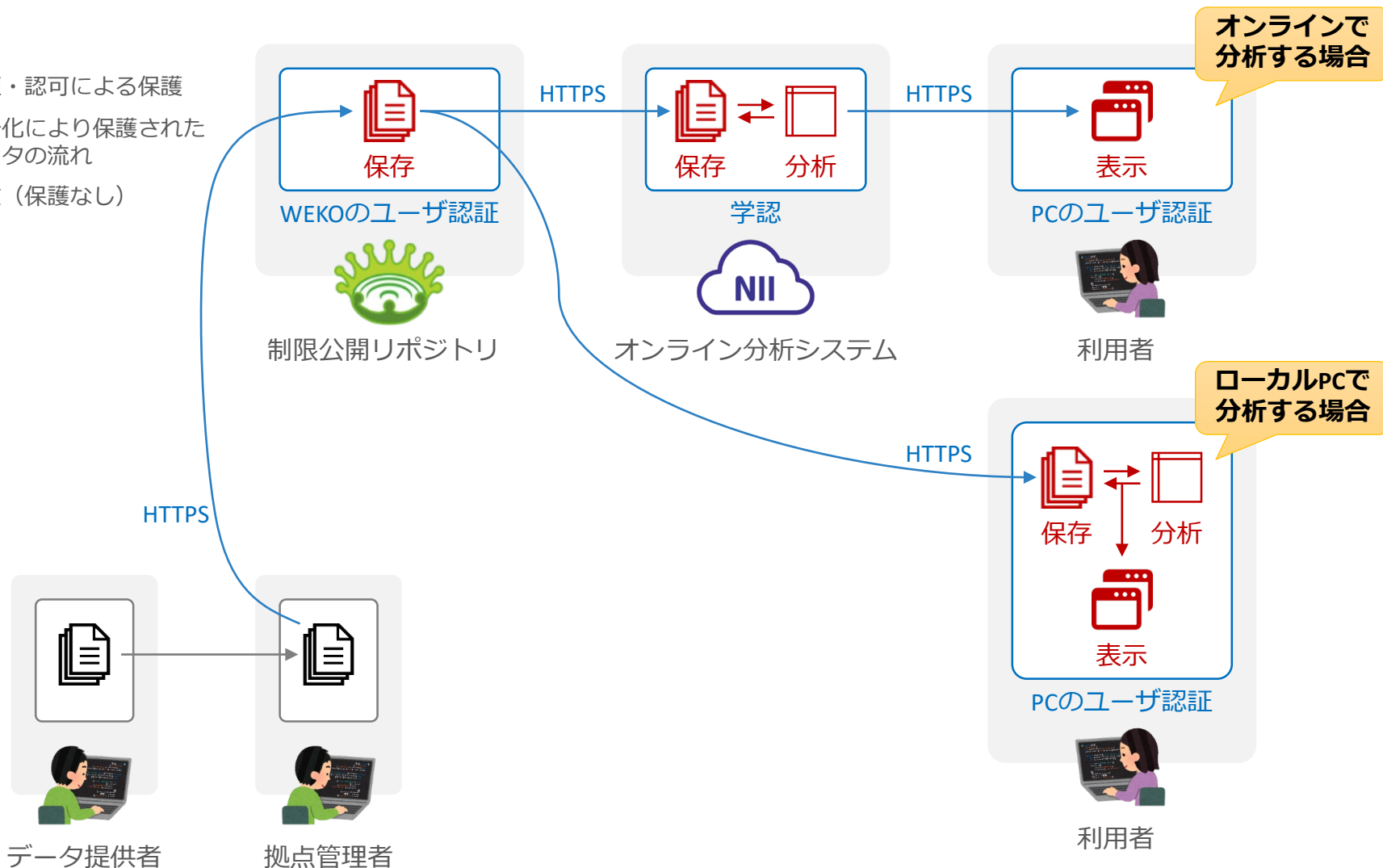
169. 人車軌道 (内訳)

年次	会社数	乗客		乗車		乗客	乗車
		乗客	乗車	乗客	乗車		
昭和1	1	20	209	2	4	16	9
2	2	20	209	2	4	16	9
3	3	20	209	2	4	16	9
4	4	20	209	2	4	16	9
5	5	20	209	2	4	16	9
6	6	20	209	2	4	16	9
7	7	20	209	2	4	16	9
8	8	20	209	2	4	16	9
9	9	20	209	2	4	16	9
10	10	20	209	2	4	16	9
計
増減

制限公開データの保護

凡例

- 認証・認可による保護
- 暗号化により保護されたデータの流れ
- 平文（保護なし）



GakuNin RDMとの連携



GakuNin RDM との連携（開発中）

GakuNin RDM マイプロジェクト 検索 Ikki Fujiwara

JH統合 ファイル Wiki メンバー **解析** 設定 証跡管理

このプロジェクトから作成した私の解析環境

解析環境名	URL	最終更新	起動	削除	🔄
vdqrb-osfstorage-6zq3agwc	/user/test001@nii.ac.jp/vdqrb-osfstorage-6zq3agwc/	2021-03-29 01:13 PM			
vdqrb-osfstorage-fp0c2a6t	/user/test001@nii.ac.jp/vdqrb-osfstorage-fp0c2a6t/	2021-03-29 01:13 PM			

新しい解析環境

このプロジェクトから作成 **外部リポジトリから作成**

環境

R (4.0.3, RStudio 1.2.5019) ✓
R is a language and environment for statistical computing and graphics. RStudio is an integrated development environment for R.

変更

① ベース環境を選ぶ

追加パッケージ

- apt-get sl: ✕ + 追加
- conda + 追加
- R (CRAN) + 追加
- R (GitHub) + 追加

② 追加パッケージを選ぶ

インストール後に実行するスクリプト

デフォルトストレージ内にpostInstallスクリプトを定義すると、イメージビルド時に任意の処理を実行できます。 [ファイル一覧を開く](#)

③ 起動！

🔄 更新

ビルド状況



GakuNin RDM との連携（開発中）

jupyterhub index Last Checkpoint: 2分前 (unsaved changes) Logout Control Panel

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3

Run Download Upload to GakuNin RDM

Setup our imports

```
In [3]: from numpy import random
        from scipy.ndimage.filters import gaussian_filter
```

Make some plots!

```
In [3]: x = random.randn(10, 500)
        x = gaussian_filter(x, [0, 10])
        sns.tsplot(x, err_style='unit_traces');
```

```
In [4]: mat = random.randn(10, 20)
        sns.heatmap(mat);
```

⑤書き戻し

④プログラム編集・実行

まとめ（再）

- 国立情報学研究所 (NII) は、JSPS人文学・社会科学データインフラストラクチャー構築推進事業の一環として「**オンライン分析システム**」を開発しています。
- **R**や**Python**で任意のプログラムが書ける分析環境を提供します。
- 学術認証フェデレーション「**学認**」の参加機関に所属する教職員や学生が利用できます。
 - 参加機関（大学等）の情報基盤担当部署で初期設定が必要です。
- JDCatで公開されているデータを1クリックで取り込み、すぐに分析できるようになります。
- 2022年度～本運用開始の予定で、現在開発中です。
- **先行試用にご協力いただける方を募集しています！**

NII オンライン分析システム

検索



RCOS
cs-support@nii.ac.jp