

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

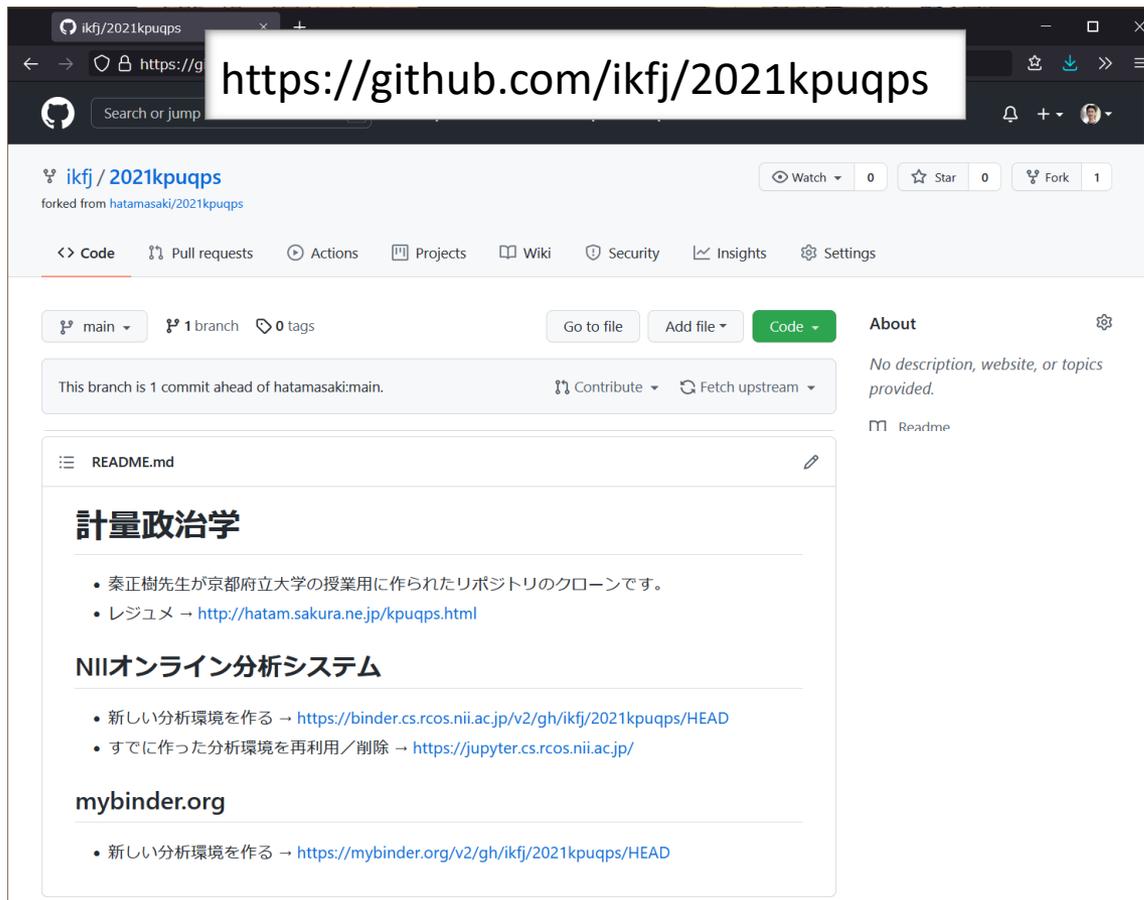


授業での先行試用

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-keiryu
			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-keiryu
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

This branch is 1 commit ahead of hatamasaki:main.

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: Chrome, URL: 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



拠点リポジトリ

GakuNin RDM

JDCatとの連携（開発中）

The screenshot shows the JDCat website interface. At the top, there is a navigation bar with 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'. Below this, there is a sidebar with filters 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'. A table of metadata is shown, including 'Item type', '公開日', 'タイトル', '作成者', '配布者', '言語', '配布者URI', '編集者', and 'URI'. On the right side, there is a '分析' (Analysis) button, a 'Versions' section showing 'Ver.1' with a timestamp, and an 'Export' section with options for OAI-PMH, JPCOAR, DublinCore, and DDI.

言語: 日本語

人文学・社会科学総合データカタログ
JDCat Japan Data Catalog for the Humanities and Social Sciences

データの言語 >

配布者 >

対象地域 >

トピック >

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

https://jdcatalog.jsps.go.jp/records/7533

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

3
views

分析

Versions

Ver.1 2021-07-09
18:28:20.796921

Show All versions

Export

OAI-PMH JPCOAR
OAI-PMH DublinCore
OAI-PMH DDI

JDCatとの連携 (開発中)

The screenshot shows a Jupyter Notebook environment with the following content:

```
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]:

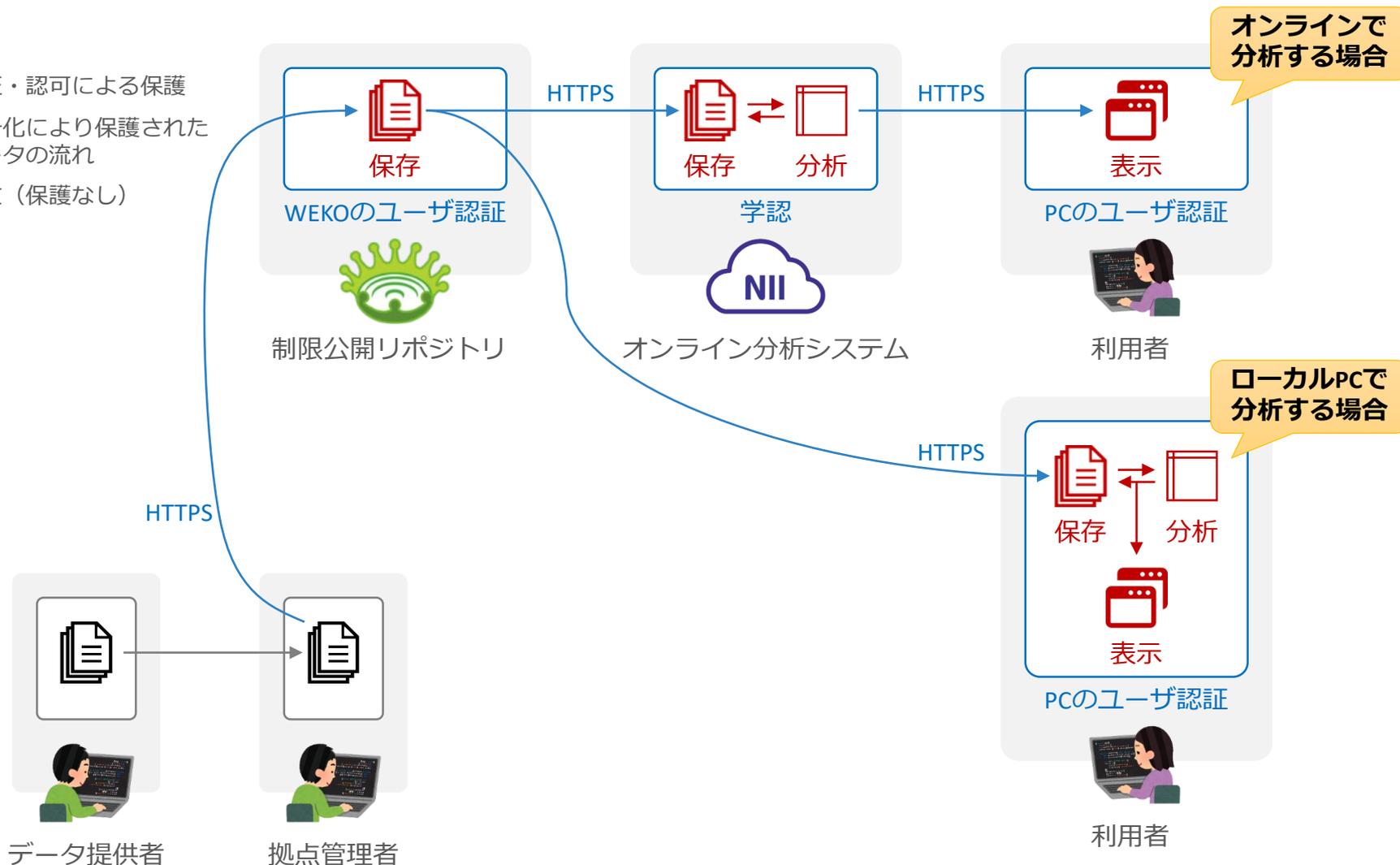
The previewed PDF document contains several tables of data, including:

- 163. 鉄道営業収支 (単位:千円)**: A table with columns for '種別' (Category) and '種別標準' (Standard Category), and rows for '収入' (Income) and '支出' (Expense) from 1937 to 1949.
- 164. 地方鉄道概況 (内訳)**: A table with columns for '年度' (Year), '線区' (Line/Zone), '営業キロ' (Operating km), '乗客' (Passengers), and '貨物' (Freight).
- 165. 鉄道事故**: A table with columns for '年度' (Year), '種別' (Category), '件数' (Number of incidents), and '死傷者数' (Number of deaths/injuries).
- 166. 電気軌道 (内訳)**: A table with columns for '年度' (Year), '種別' (Category), '乗客' (Passengers), '乗車' (Ridership), '人員' (Staff), and '設備' (Equipment).
- 167. 汽船軌道 (内訳)**: A table with columns for '年度' (Year), '種別' (Category), '乗客' (Passengers), '乗車' (Ridership), '人員' (Staff), and '設備' (Equipment).
- 168. 馬車軌道 (内訳)**: A table with columns for '年度' (Year), '種別' (Category), '乗客' (Passengers), '乗車' (Ridership), '人員' (Staff), and '設備' (Equipment).
- 169. 人車軌道 (内訳)**: A table with columns for '年度' (Year), '種別' (Category), '乗客' (Passengers), '乗車' (Ridership), '人員' (Staff), and '設備' (Equipment).

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）

The screenshot shows a JupyterLab environment with the following elements:

- Header:** jupyterhub index Last Checkpoint: 2分前 (unsaved changes) | Logout | Control Panel
- Menu:** File | Edit | View | Insert | Cell | Kernel | Widgets | Help
- Toolbar:** Not Trusted | Python 3 | Run | Download | Upload to GakuNin RDM
- Section 1: Setup our imports**

```
In [3]: ▶ from numpy import random
        from scipy.ndimage.filters import gaussian_filter
```
- Section 2: Make some plots!**

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

```
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