

2024年3月12日,14日,19日,22日

オンライン開催

研究データに関わる
研究・教育支援人材向け

Python 活用基礎研修（導入）

附属図書館 研究開発室

甲斐 尚人・神崎 隼人

主催：附属図書館研究開発室

共催：産業科学研究所 AI センター

コアファシリティ機構

(データ利活用 DX 支援部門・研究支援人材育成部門)

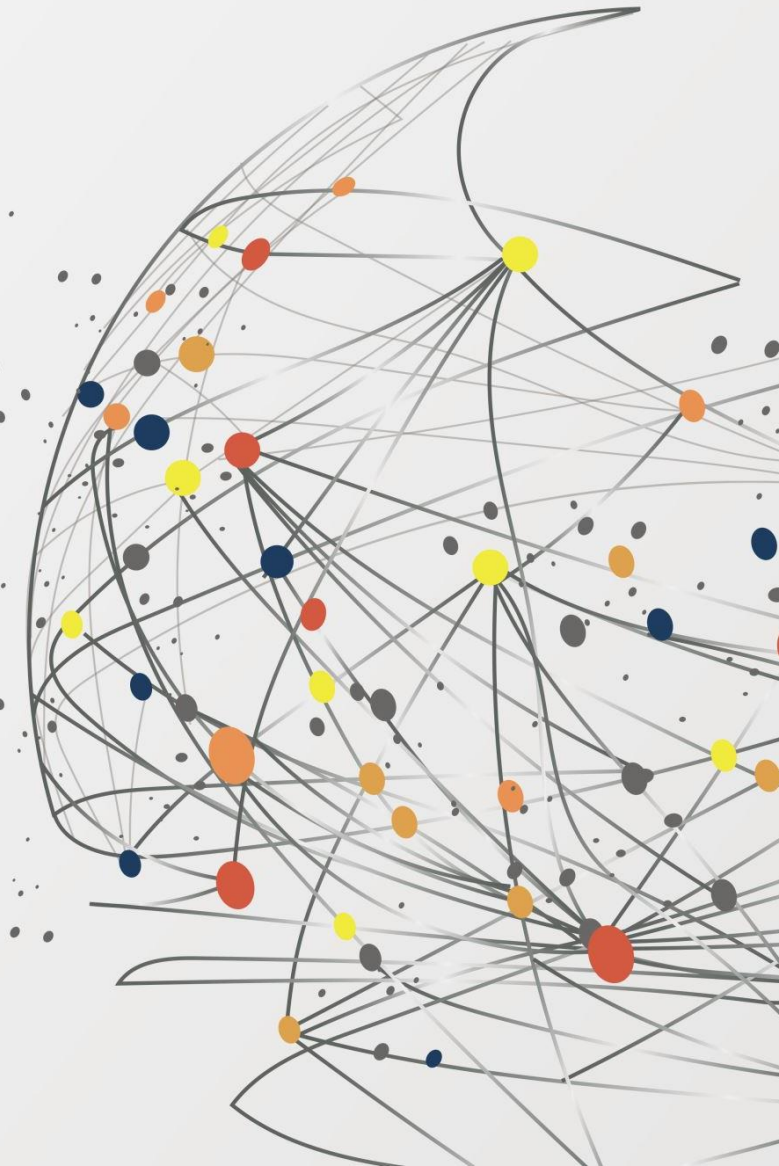
情報推進部

サイバーメディアセンター

協力：技術職員連絡会議

部局横断型女性技術職員ネットワーク

産業科学研究所技術室



研究データの管理の観点から

大阪大学研究データポリシー

3. 研究者の責務

3.2 研究データの公開・利活用

～中略～また、公開にあたっては、利活用を促進するために研究データの品質の確保に努めなければならない。

研究データポリシー学内説明会資料より

大阪大学
OSAKA UNIVERSITY



研究データの取得・収集、保存

当該研究者が判断する適切な方法により、研究データを取得・収集し、保存する

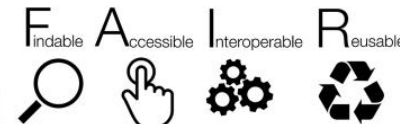
- ✓ 関係する法令、学内規則、研究倫理その他の規範の遵守
- ✓ データ来歴を明らかにした上で保存
- ✓ 保存期間終了後には適切に処理

研究活動の公正性
研究データの価値向上

研究データの公開・利活用

研究者の判断の上、可能な限り当該データを社会に公開し、その利活用の促進に努める

- ✓ FAIR原則
データの共有・公開時に考慮すべき原則
ただし
- ✓ 関連する法令等
- ✓ オープン・アンド・クローズ戦略（前出）



出典：FORCE11: THE FAIR DATA PRINCIPLES (2016).



公開段階では研究データの品質の確保に努める

データの信頼性 相互運用性 正確性 機械可読性 トレーサビリティなど

11

✓ クリーニングされたデータと適切なデータ管理は、研究の再現性とデータの再利用性を確保

★ 大量・大容量データの正確で効率的な処理が必要

研究データの品質確保（クリーニング・処理）



Types of Error

Standardization and Formatting (spaces, dates, strings, ...)

Missing values (empty cells, unwanted rows, ...)

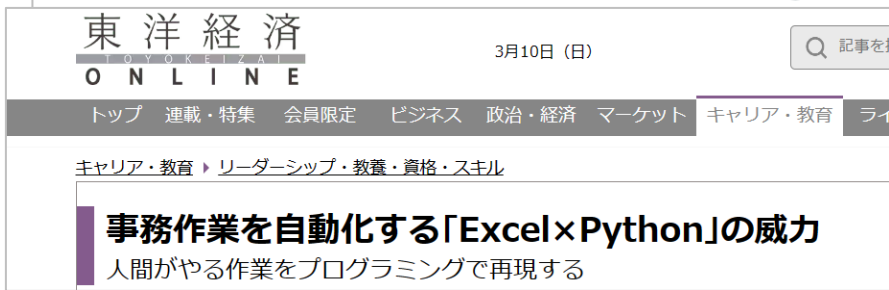
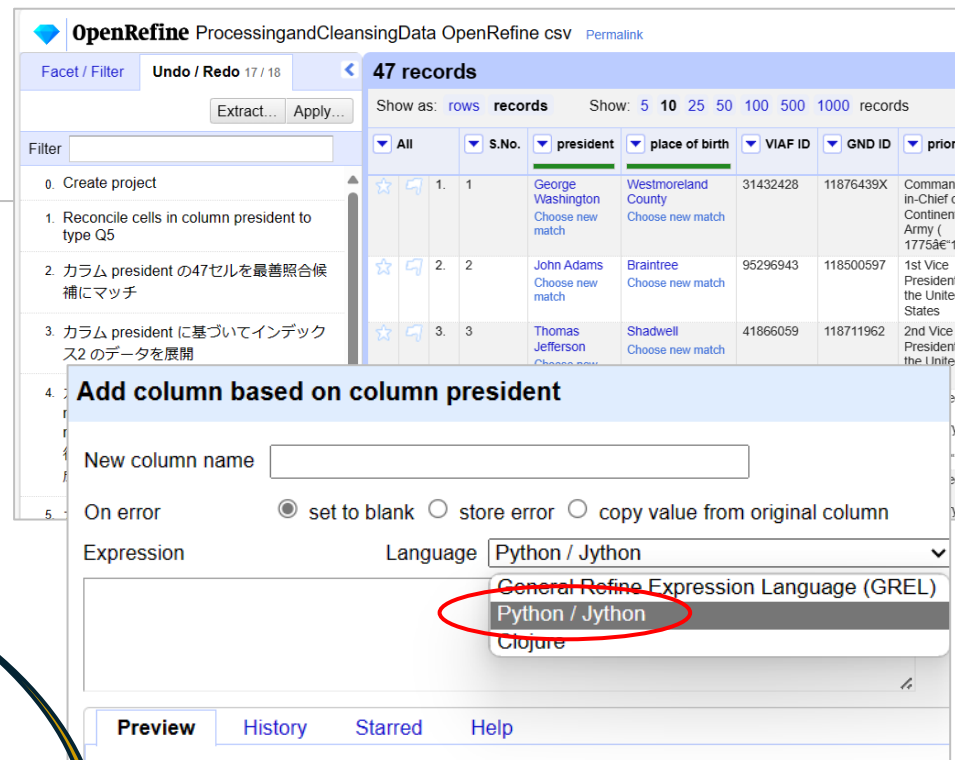
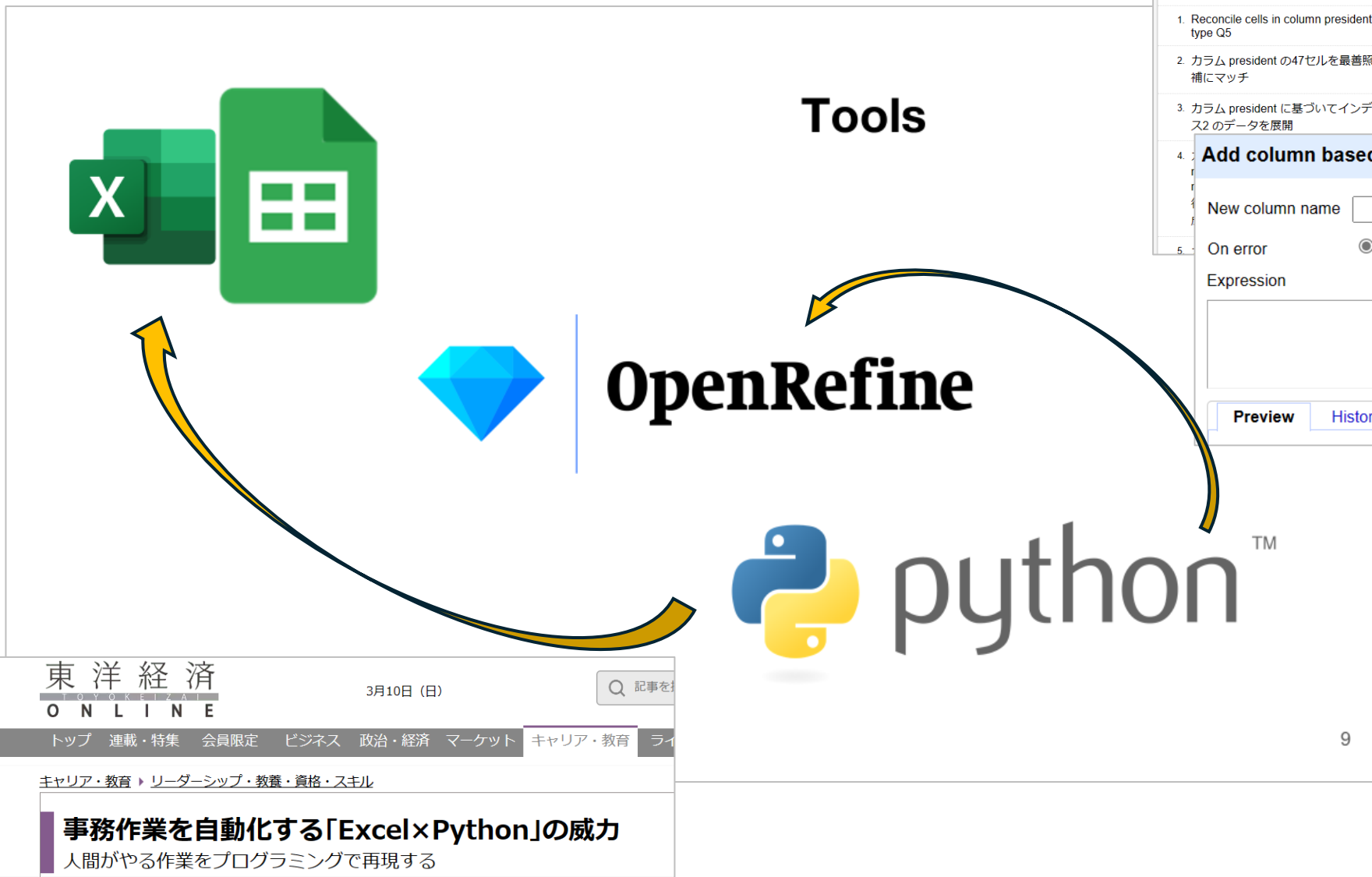
Bad values (person is 200 years old, the price of a carrot is 280 €, ...)

Duplicates

Invalid characters and different encoding (“ß” → “â-€”)

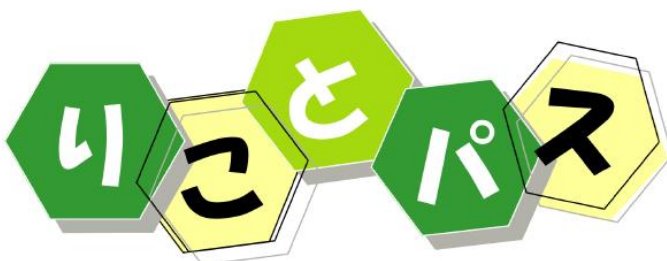
Invalid data according to a schema

データクリーニング・処理ツール

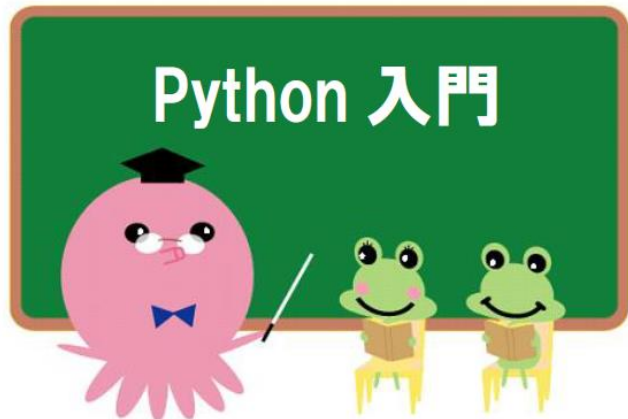


附属図書館の過去の取り組み例

No.51 (2020)



理工学図書館パスファインダー



りことパスは、主に理工学分野の授業に関連するトピックについて、学習の初めの一歩となる資料や Web サイトを紹介するテーマ別調べ方ガイドです。理工学図書館のラーニング・サポーター(LS)が作成しています。学習やレポート作成にぜひ活用してください。

2020年度 教員監修済

1. イントロダクション

プログラミングについて調べると、必ずと言っていいほど見かけるようになった Python (パイソン) についての資料を紹介します。Python はプログラミング初心者が独学でも十分マスターできるほど、簡単に扱えるプログラミング言語です。また、統計や信号処理などのデータ解析や Web アプリケーションなど、幅広い使い方ができます。実は、YouTube や Instagram も Python で開発されています。「とりあえずプログラミングのスキルを身に付けたい」という方は是非 Python を勉強してみてください!!

1-1. Python の特徴

いざプログラミングを学ぼうと調べてみると、多種多様なプログラミング言語 (C, C++, Java, Ruby, JavaScript, Go, Swift, ...) があることに気づくと思います。これだけあるとどれを触ればいいのか迷ってしまいますよね。その中で Python が初心者に人気な理由は、「学びやすさ」にあると思います。一般的に Python が「文法がシンプル」「コードが読みやすい」と言われていたことからも窺い知ることができると思います。

以下の Web サイトでは、簡単に Python の特徴を説明しているのて、勉強をする前にまず Python について知りたい方は、一度目を通してみてください。

- YouTube や Instagram を開発! 人気急上昇中のプログラミング言語、Python とは / INTERNET ACADEMY -IT まるわかりガイド- (2019. 3. 29 掲載)
https://www.internetacademy.jp/it/programming/ei/what-is-python_the-popular-programming-language.html
- プログラミング言語 比較鳥瞰図 / Mimir Yokohama (2020. 2. 18 更新)
<https://www.mimir.yokohama/useful/0005-proglang-comp.html>

様々なプログラミング言語の特徴を簡単に短く紹介しています。簡単なサンプルコードと共に紹介しており、分かりやすいです。

2. Python 入門書

はじめてプログラミングを学ぶ方は、ここで紹介する本から読んでみてください。パソコンへの Python のインストール、プログラミングとは何かなど、初学者向けの内容を含んでいる本を選びました。





2-1. 入門書

- ゼロからわかる Python 超入門 / 佐藤美登利著, 2018 年

【書誌 ID=2004490695】

コンピュータの仕組みにはじまり、Python のインストールの仕方から Python のことを教えてくれます。また Atom と呼ばれるエディタのインストールと日本語化までサポートしてくれます。ただし、Windows を基準にインストールの説明がされているので、Mac・Linux ユーザーは注意が必要です。説明がとても詳しく、サンプルコードや例題もあるため、この本にあるコードを写経(書き写し)し

scikit-learn を
に關しては、Mac・
ID=2004467767】
ータ分析用ライブ
す。これらのライ
になります。
ID=2004484929】
いませんか?
の性能を發揮する
習を行う予定の方
ID=2004487195】
ます。実はこの
題材にその使い
python に計算を
らがオススメです。
りごたえのある
年
ID=2004490441】
りがオススメです。
りごたえのある
別)に多くの参考書
これらの参考書を紹
学習を目的として
きる本です。他に
して向いています。
2017 年
ID=2004489989】
脱初心者を目指し
に関する内容も追加
み始めるといい
ID=2004461374】
oj)を使いながら、
ったことがない方
り入門書で、Python
んでから、この本



Twitter

Facebook

1

2

3

4

今回Python活用基礎研修を実施

■ 産業科学研究所AIセンター 木村先生・山中技術専門職員 作成教材より

Python活用基礎研修について

研修の目標

- Pythonを通してプログラミングの基礎を学ぶ.
- 実際に手を動かして簡単なデータ解析およびAI（機械学習）を体験する.

入門編

- 入門編1: Pythonプログラミングの基本的なルールを学ぶ.
- 入門編2: NumPyを用いたプログラミングを学ぶ.

応用編: 各種モジュール（=便利なツール）の使い方を学ぶ

- 応用編1（データ解析）: pandas, Matplotlib
- 応用編2（AI, 機械学習）: scikit-learn

主催： 附属図書館研究開発室
共催： 産業科学研究所 AI センター
コアファシリティ機構
(データ利活用 DX 支援部門・研究支援人材育成部門)
情報推進部
サイバーメディアセンター
協力： 技術職員連絡会議
部局横断型女性技術職員ネットワーク
産業科学研究所技術室

8. 講師

講師： 木村 輔 特任助教（常勤）（産業科学研究所・コアファシリティ機構）
山中 卓也 技術専門職員（産業科学研究所 AI センター）

講師補佐：光岡 孝 特任研究員（産業科学研究所）・奥村 由香 係長（産業科学研究所）
川妻 恵理 技術専門職員（工学研究科）・向園 悠特任技術職員（情報推進部）
甲斐 尚人 准教授（附属図書館）

研究推進部や附属図書館の取り組み 研究データ管理教材の紹介

The screenshot shows a video player interface for a course titled "研究データマネジメントを知る" (Learn about Research Data Management). The interface includes a navigation bar at the top with icons for book, topic, link, live broadcast, learning analysis, and tag management. The course is by Kai Naoto (甲斐 尚人), an instructor (LTI情報). The video duration is 9 minutes and 49 seconds. A table of contents is displayed on the right, listing 8 topics. The first topic, "オープンサイエンス時代における研究データ管理の基礎" (Basics of Research Data Management in the Open Science Era), is currently selected and has a duration of 1 minute and 10 seconds. Below the table of contents, there is a "タグを追加" (Add tags) button and a "解説" (Explanation) section with a "トピックの詳細" (Topic details) link. The main content area shows the beginning of the video, with the text: "本教材では、新任教員研修プログラムとして、オープンサイエンス時代における研究データマネジメントの基礎について学ぶことができます。" (In this material, you can learn the basics of research data management in the open science era as part of the new faculty training program.)

viLOG ブック トピック リンク 配信中のブック 学習分析 タグ管理 Kai Naoto 甲斐 尚人 教員 LTI情報

研究データマネジメントを知る

学習時間 9分 49秒 他ブックを配信

番号	タイトル	学習時間
1	オープンサイエンス時代における研究データ管理の基礎	1分 10秒
2	研究データマネジメントの必要性	1分 34秒
3	研究データマネジメントの意義	37秒
4	研究データのライフサイクル	2分 7秒
5	世界と日本におけるオープンサイエンスの現状	2分 40秒
6	研究者の負担軽減のための取り組み	1分 41秒

目次

1. 研究データマネジメントを知る
2. 研究データポリシーと関係規定等
3. 研究データ管理計画
4. 研究データを取得・収集および保存・共有する
5. 研究データを公開する
6. 大阪大学機関リポジトリ OUKA
7. 研究データを利活用する
8. 大阪大学の研究データ基盤・連携の紹介

タグを追加

解説 トピックの詳細

本教材では、新任教員研修プログラムとして、オープンサイエンス時代における研究データマネジメントの基礎について学ぶことができます。

本教材を提供するオープンサイエンス推進室は、研究データポリシーの策定やその改訂をはじめ、学内



それでは

Python活用基礎研修の始まりです

講師の木村先生、山中様（技術専門職員）お願いいたします