Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection Library | SMU Libraries

11-2022

The emerging area of data services: A case of SMU Libraries

Bella RATMELIA
Singapore Management University, bellar@smu.edu.sg

Follow this and additional works at: https://ink.library.smu.edu.sg/library_research

Part of the Asian Studies Commons, and the Library and Information Science Commons

Citation

RATMELIA, Bella. The emerging area of data services: A case of SMU Libraries. (2022). *13th International CALIBER 2022, India, November 17-19*. 1-28.
Available at: https://ink.library.smu.edu.sg/library_research/204

This Conference Paper is brought to you for free and open access by the SMU Libraries at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Library by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

SMU SINGAPORE MANAGEMENT UNIVERSITY | Libraries

The emerging area of Data Services: A case of SMU Libraries

Bella Ratmelia (bellar@smu.edu.sg)

Data Services Librarian, SMU Libraries

18 November 2022

Kwa Geok Choo Library (Law Library)

Li Ka Shing Library (Main Library)

SMU
STAMFORD HOUSE

About SMU

6 schools
2 colleges
13,000+ students
388 faculty from 26 nationalities
20+ research institutes/centers

SMU's Expansion

- Creation of new colleges with focus on inter-disciplinary research:
  - College of Graduate Research
  - College of Integrative Studies
- Creation of new programs with emphasis on data skills:
  - MSc in Accounting (Data & Analytics)
  - 2nd Major in Accounting Data & Analytics
  - 2nd Major in Financial Forensics (2020)
  - Bachelor of Science (Computing & Law)

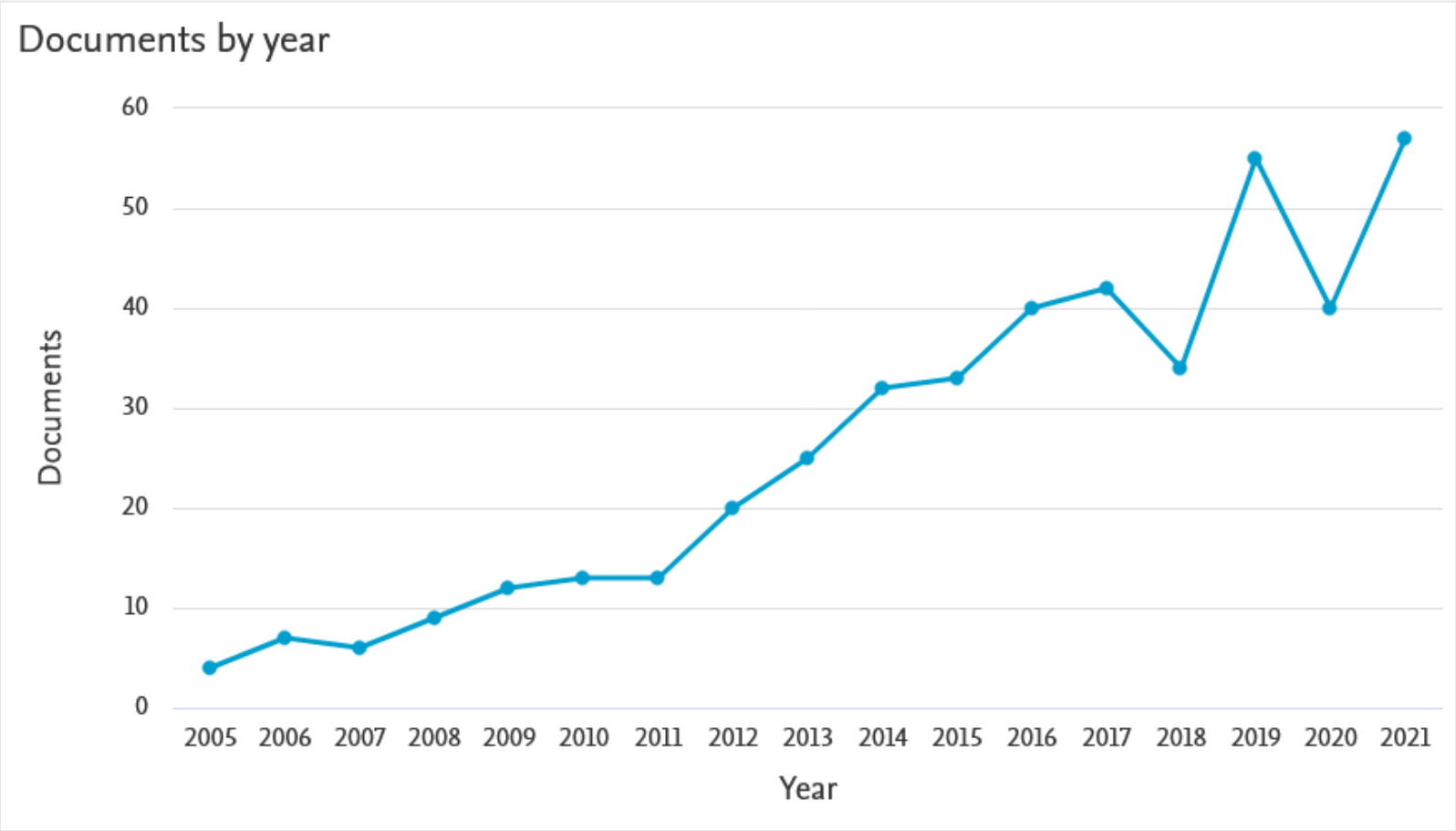
Questions we get now…

“I have gathered about 1 million articles on topic A and B. Do you have any python code snippets that I can use for a simple text analysis e.g., word count?”

“I need to analyse all news about company X, Y, Z. I found that there are around 25,000+ news in total in Factiva. How do I get retrieve them? I heard you can do this via an API? I am really a rookie and have no experience on this before. I hope you can share some of your experience to help me learn and work on this.“

"I noticed <tool> allows you to extract references in the bibliography from PDFs. I have been looking for a tool to do this. Are you aware of anything that does this?"

“I am using UN Comtrade API to retrieve financial data for my research. But I keep having trouble with it. Can you help?”



Number of TDM-related papers by SMU Faculty every year from 2005 to 2021

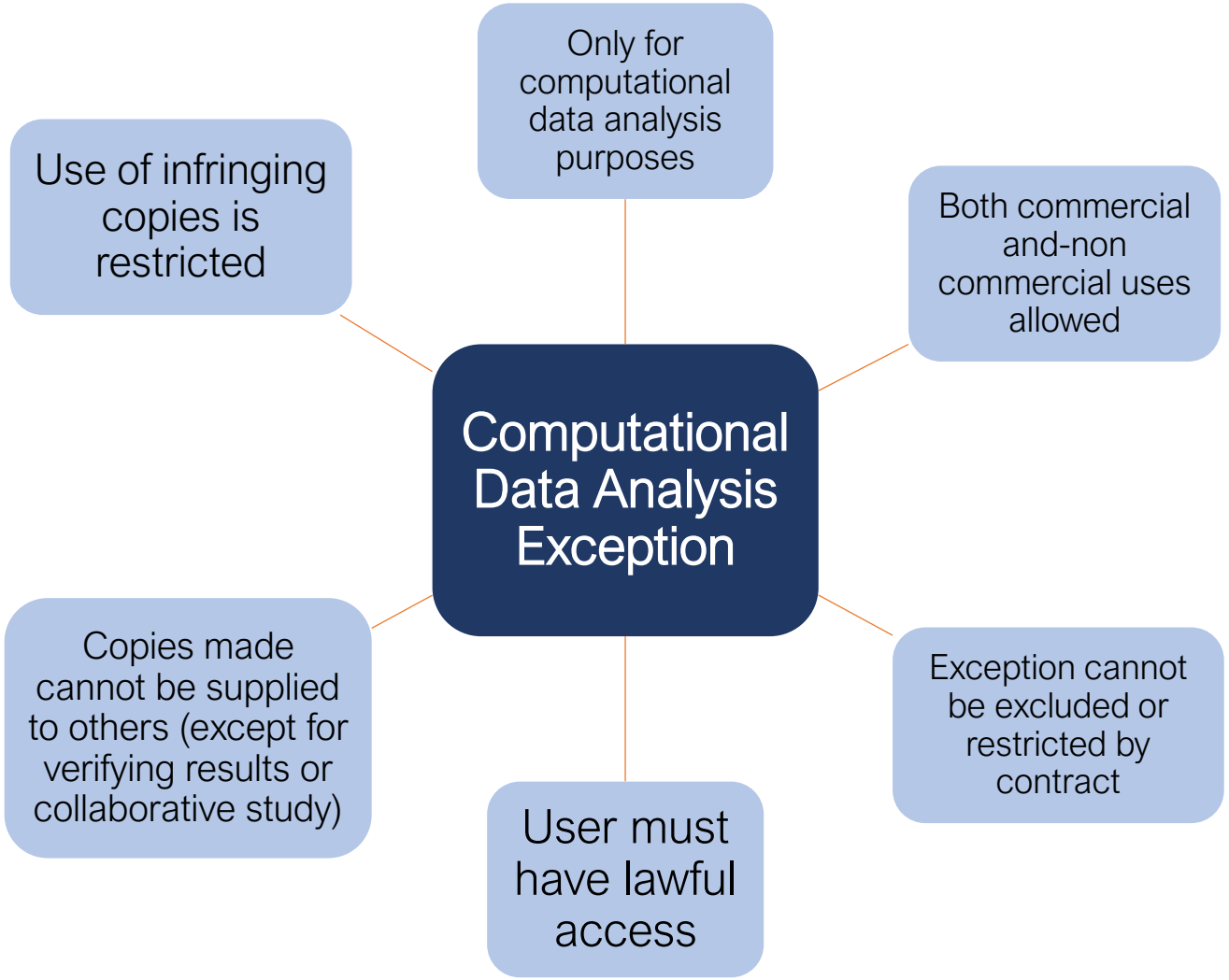
A quick search on Scopus revealed that there are approximately **442** TDM-related papers by SMU faculty since 2005 to 2021.

The numbers are increasing every year, and this trend is unlikely to subside soon.

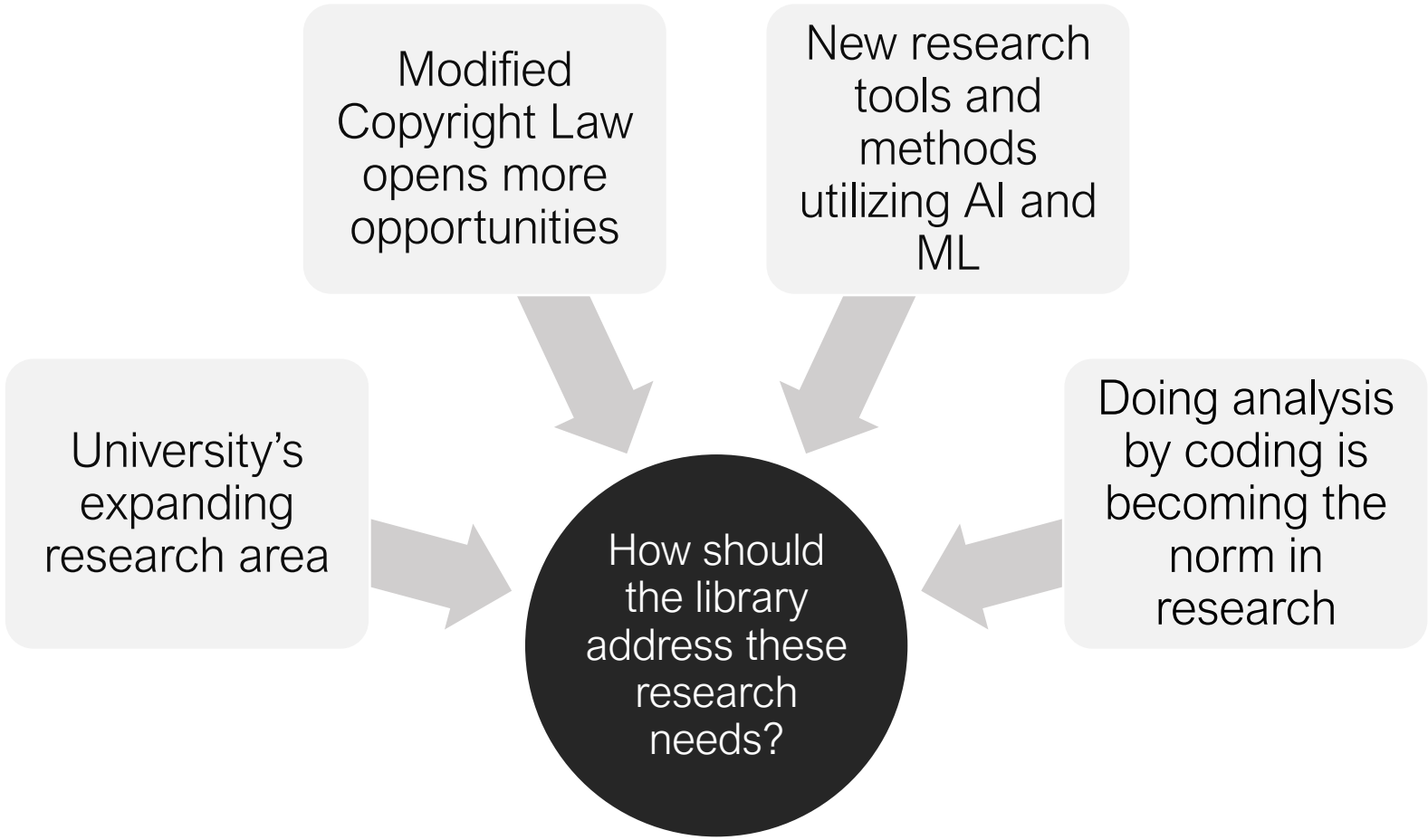
New Exception in SG Copyright Law:

Computational Data Analysis Exception (CDA)

- Took effect on November 2021
- Now equipped with advantageous features!



The issue summarized:



SMU SINGAPORE MANAGEMENT UNIVERSITY | Libraries

Our Response

Answers to challenges and changes

Adapted from "Innovations in Scholarly Communication" by Kramer & Bosman

SMU SINGAPORE MANAGEMENT UNIVERSITY | Libraries

What exactly is “Data Services?”

One possible current definition:

“Services that address the full data lifecycle, including the data management plan, digital curation (selection, preservation, maintenance, and archiving), and metadata creation and conversion.”

(Tenopir et al., 2012)

What exactly is “Data Services?”

But we define it as such:

“a suite of services that support and enable the use of computational methods in research.”

Approaches

1

Building the human capital

2

Enabling the resources (and services)

3

Engaging & connecting users to resources

4

Keeping abreast with latest developments

1

Building the Human Capital

Get a dedicated librarian to develop and deliver data services

- More investment (new headcount)
- More dedicated time to develop skills and knowledge required
- Easier to scale up services

VS

Train existing staff to develop and deliver data services

- Cheaper as it uses existing headcounts
- Less time to develop skills and knowledge required
- Scaling up services won't be as easy

Why not both?

Get a dedicated librarian and have them train existing staff!

SMU SINGAPORE MANAGEMENT UNIVERSITY | Libraries

Building the Human Capital

Dedicated Data Services Librarian

To train existing staff, develop new services and expand existing ones, etc.

Library Carpentry for Library Staff

2 days workshop on Python, OpenRefine, and Tidy Data.

Bi-monthly RDS Session

1 hour discussion on topics in research e.g., bibliometrics, software citations, data tools, etc.

Class Audit

Audit relevant classes to keep up with the situation “on the ground”

Keep up with latest developments right from the “source”.

2

Enabling the resources and services

From easiest to hardest:

- Creating LibGuides and FAQs based on online resources evaluation (free and paid)
  - Including working with users to do these evaluations
- Providing relevant workshops
- Providing the physical space/resources
- Outreach and promotion

2

List of APIs for Scholarly Resources

A list of commonly used scholarly resources API

Search this Guide | Search

- Home
  - Introduction
  - FAQs
  - Check out other useful guides below
- Scholarly Metadata API
- Elsevier API (ScienceDirect & Scopus)
- Web of Science API
- Other Data APIs
- Social Media APIs

Data Services Librarian

Introduction

What is an API?

API stands for **Application Programming Interface**. Essentially, it's a standardized method that enables online software/services to communicate with each other by way of exchanging data. APIs are used in many different ways such as adding new data on a website through another (e.g. automatically tweet a new blog post), retrieve data from publicly open data repositories, or update contents that resides in another website.

In research and academia, however, data retrieval is the most common usage of API. Generally, you will need some knowledge on programming (e.g. R or Python) to use API.

What is this API guide about?

This research guide lists some of the well-known and often-used APIs in research and academia. Simply use the tabs on the left to browse the list.

Questions or Feedback?

If you have any feedback, questions, or know of other APIs that you think should be included in this list, email us at library@smu.edu.sg.

FAQs

How do I find or search within the 50 Journals used in the FT Research Rank (FT50 Journals)?

SMU SINGAPORE MANAGEMENT UNIVERSITY | Libraries

2

Analysis and Visualization Tools

Learn how to do your research faster and better!

- Introduction
- 1. Preparation
- 2. Discovery
- 3. Analysis and Visualization
  - Activities in this Phase
  - Data Wrangling Tools
  - Statistical Analysis Tools
  - Text Analysis Tools
  - Bibliometric Analysis Tools
  - Business Intelligence Tools
- 4. Writing
- 5. Publication
- 6. Outreach
- 7. Assessment

3. Analysis

Common research activities in this phase include **data collection** (through existing datasets, first-hand collection, or experiments), **data analysis**, and **documenting/maintaining the protocols or instruments** used in the analysis.

Data Wrangling Tools

Tools with no coding required (point-and-click)

- OpenRefine
  OpenRefine is an open source application used for data wrangling (data cleaning, data transformation, data parsing, etc.). You can also use the application to fetch additional data from API endpoints such as Crossref API and many others.
- Power Query
  Power Query is a tool for data preparation that is integrated by default in Excel (Office 365 or Excel 2016) or Power BI. It enables you to connect to external data sources such as API or Sharepoint sites, and combine, reshape, and manipulate the data. Official documentation of Power Query

Tools with coding required

- RStudio
  RStudio is an open-source integrated development environment (IDE) for R. It comes with debugging tools, syntax highlighters, and other features that makes working with R easier and more manageable. Download and installation instruction here.

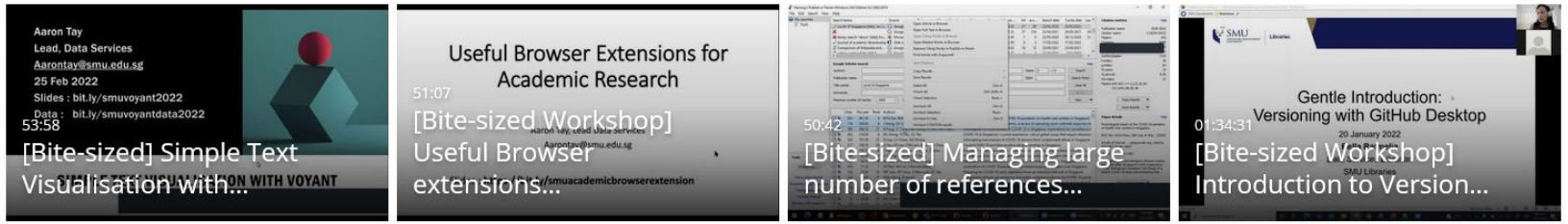
SMU SINGAPORE MANAGEMENT UNIVERSITY | Libraries

2 3

Workshops on Python, R, and other useful research data tools

- Python workshops consistently in top 5 most popular workshops.
  - ~180 attendees is the highest number so far
- Average attendance of ~50 people per session
- Workshops on literature reviews, text and data mining tools also very popular

WORKSHOPS

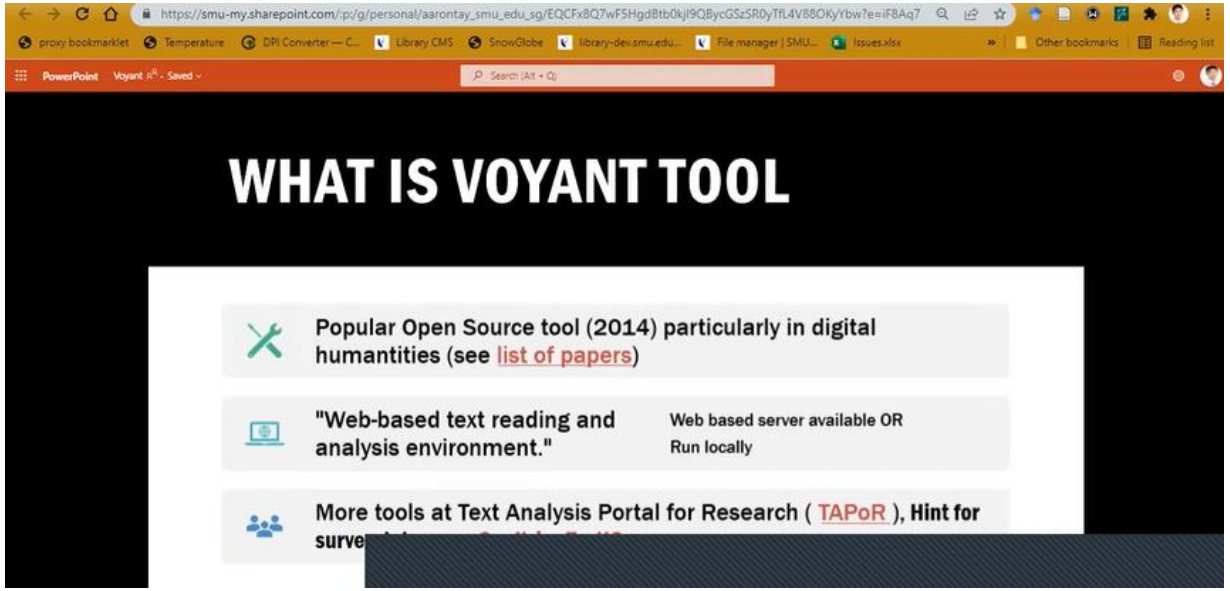


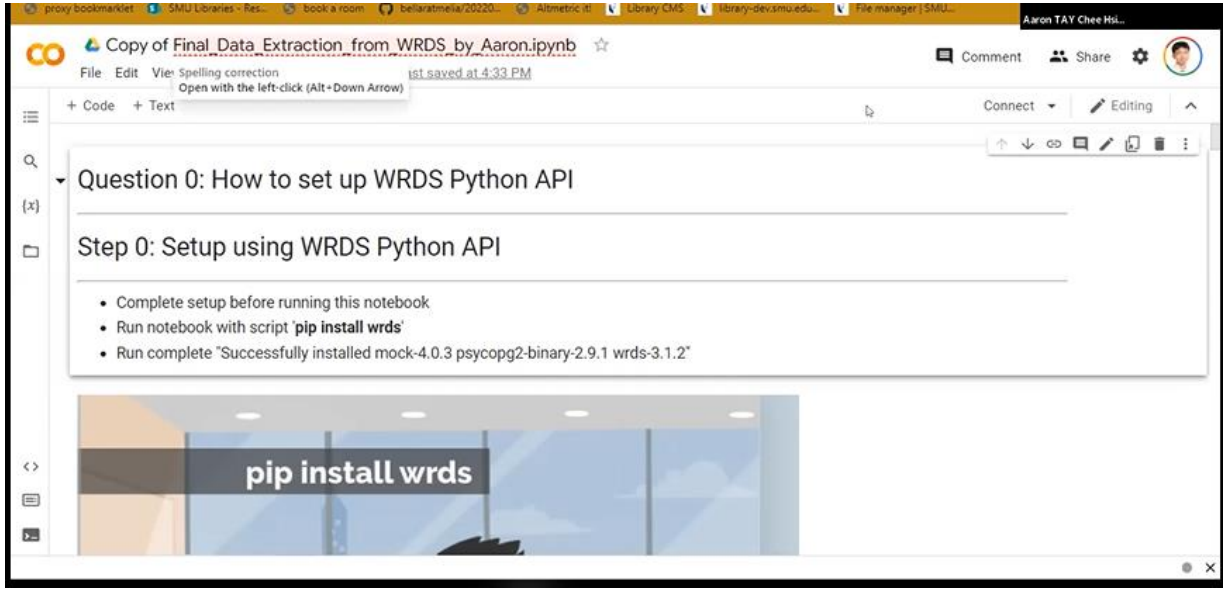
2022 January-February workshops

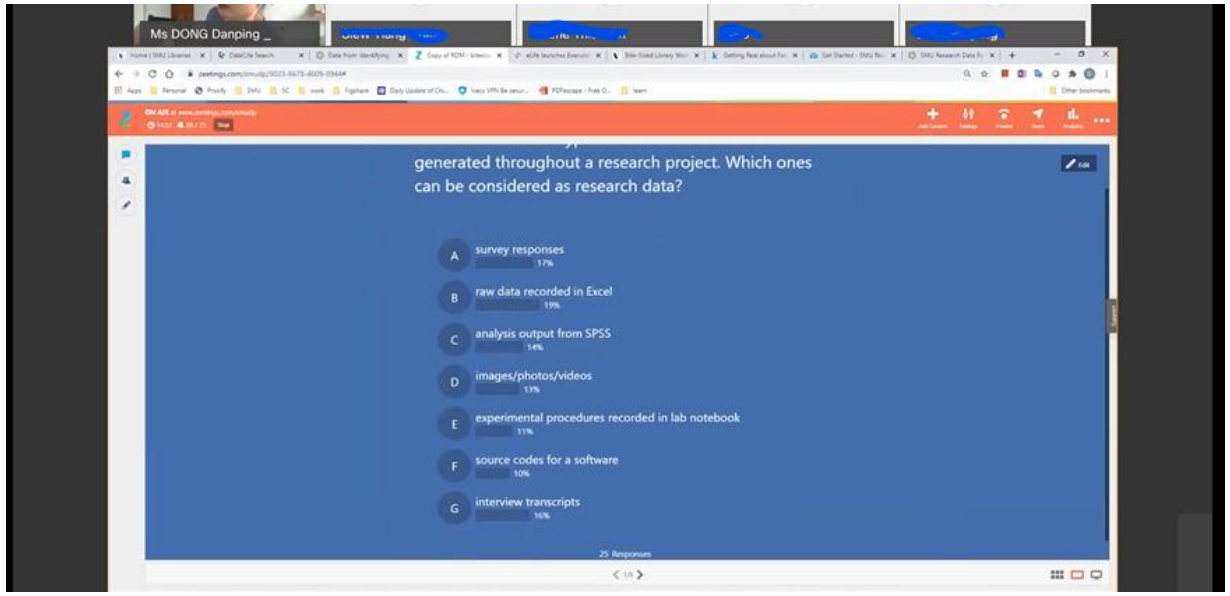
- Simple Text Visualisation with Voyant Tool. Learn how to do simple text analysis and visualization using Voyant, a light weight and free open source tool.
- Useful Browser extensions for academic research. Aaron Tay, Lead Data Services share some of his favourite Chrome browser extensions for research such as scite, LibKey Nomad, and more.
- Managing large number of references efficiently. Learn the different ways to efficiently export large number of references from sources like Google Scholar or from reference sections of PDF and Word documents. No full text available? Learn how to use tools that leverage citation indexes to extract references and citations from a set of papers of interest.
- Python 101 Part 1: The Basics. This session covers the fundamentals for novice coders, including arithmetic operation, variables, data types, list, For loops, and conditional statements (if-else)
- Python 101 Part 2: Tinkering with Data Interchange Format. This session covers the concept of Python libraries, basic concepts on Panda's DataFrame, JSON, and XML.
- Python 101 Part 3: Tinkering with Scopus API. This session covers basic concept of API and a beginner walkthrough on how to retrieve Scopus search result using the API.
- Introduction to Version Control Using GitHub Desktop. Covers the basic concepts of version control, demonstrated using GitHub Desktop (beginner friendly!)
- Going beyond Google Scholar, Scite, Overton and new discovery indexes. Learn about how Google Scholar compares to scite, a new SMU Libraries' subscription search engine that allows searching via citation statements as well as filtering citations by citation type ("mentioning", "supporting", "constrasting")

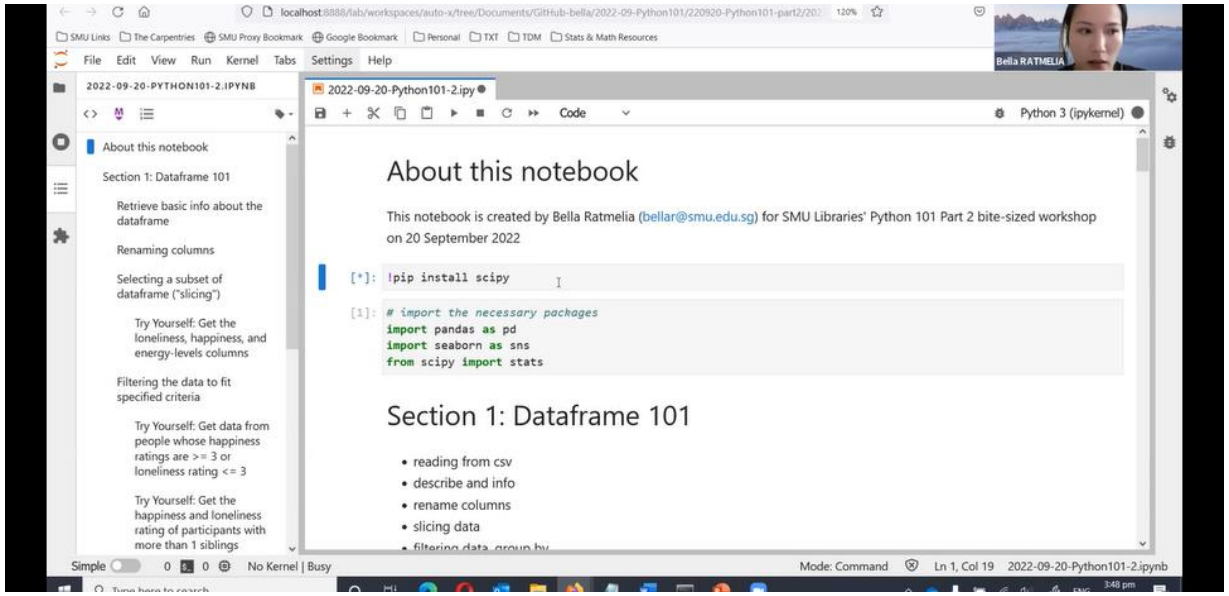
SMU SINGAPORE MANAGEMENT UNIVERSITY | Libraries

2 3


WHAT IS VOYANT TOOL
Popular Open Source tool (2014) particularly in digital humantities (see list of papers)
"Web-based text reading and analysis environment."
Web based server available OR
Run locally
More tools at Text Analysis Portal for Research (TAPoR), Hint for


Copy of Final_Data_Extraction_from_WRDS_by_Aaron.ipynb
Comment
Share
Connect
Editing
Question 0: How to set up WRDS Python API
Step 0: Setup using WRDS Python API
Complete setup before running this notebook
Run notebook with script 'pip install wrds'
Run complete "Successfully installed mock-4.0.3 psycopg2-binary-2.9.1 wrds-3.1.2"
pip install wrds


Ms DONG Danping
generated throughout a research project. Which ones can be considered as research data?
A survey responses
B raw data recorded in Excel
C analysis output from SPSS
D images/photos/videos
E experimental procedures recorded in lab notebook
F source codes for a software
G interview transcripts
25 Responses


Bella RATMELIA
2022-09-20-PYTHON101-2.IPYNB
About this notebook
Section 1: Dataframe 101
Retrieve basic info about the dataframe
Renaming columns
Selecting a subset of dataframe ("slicing")
Try Yourself: Get the loneliness, happiness, and energy-levels columns
Filtering the data to fit specified criteria
Try Yourself: Get data from people whose happiness ratings are >= 3 or loneliness rating <= 3
Try Yourself: Get the happiness and loneliness rating of participants with more than 1 siblings
2022-09-20-Python101-2.ipy
Code
Python 3 (ipykernel)
About this notebook
This notebook is created by Bella Ratmelia (bellar@smu.edu.sg) for SMU Libraries' Python 101 Part 2 bite-sized workshop on 20 September 2022
[*]: !pip install scipy
[1]: # import the necessary packages
import pandas as pd
import seaborn as sns
from scipy import stats
Section 1: Dataframe 101
reading from csv
describe and info
rename columns
slicing data
Simple
No Kernel | Busy
Mode: Command
Ln 1, Col 19
2022-09-20-Python101-2.ipynb

2 3

Physical space

(Research Data Services) RDS Studio

“Provide equitable access to equipment”

User needs assessment

- Focus groups and interviews with students and faculty

Findings & Strategy

- Start small
- Make use of existing resources
- Services > physical resources

2 3 RDS Data Studio - Provide equitable access to equipment

Monthly ResearchRadar

- Articles on interesting developments pertaining to research/data in general
- Written by librarians!
- This year, it receives **250** views per article on avg.

RESEARCHRADAR

OSF vs. RDR – where should I put my data?
Fri, 10/28/2022 - 12:00

RESEARCHRADAR

Learn basic data and library skills by reviewing recordings of Bite-Sized workshops – 2022/2023
Fri, 10/28/2022 - 12:00

RESEARCHRADAR

Creating search strategies in Scopus to match results from a list of selected journal titles (e.g., FT50 Journal List)
Fri, 10/28/2022 - 12:00

RESEARCHRADAR

Creating search strategies to match results from a list of selected journal titles (e.g., FT50 Journal List) in Web of Science
Fri, 09/30/2022 - 12:00

RESEARCHRADAR

IEEE DataPort vs. SMU RDR: Where should I put my data?
Fri, 09/30/2022 - 12:00

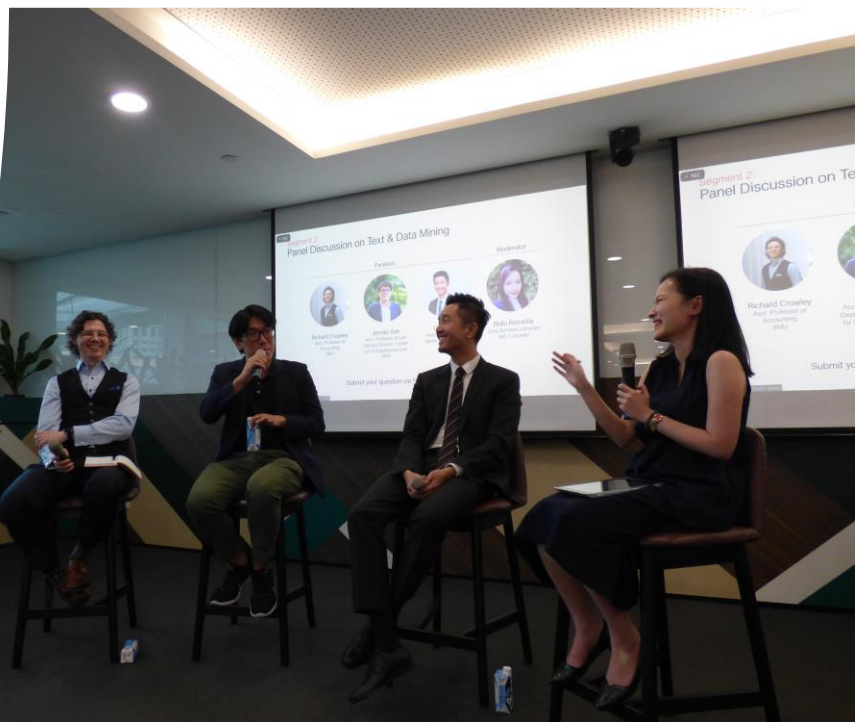
RESEARCHRADAR

The Bleeding Edge – Updates on LibKey Nomad, Mendeley Desktop, Publons profiles and the new public access mandate
Fri, 09/30/2022 - 12:00

SAUL Copyright Seminar 2022

Collaboration with other university libraries and government bodies (IPOS) to raise awareness of new changes in SG copyright law

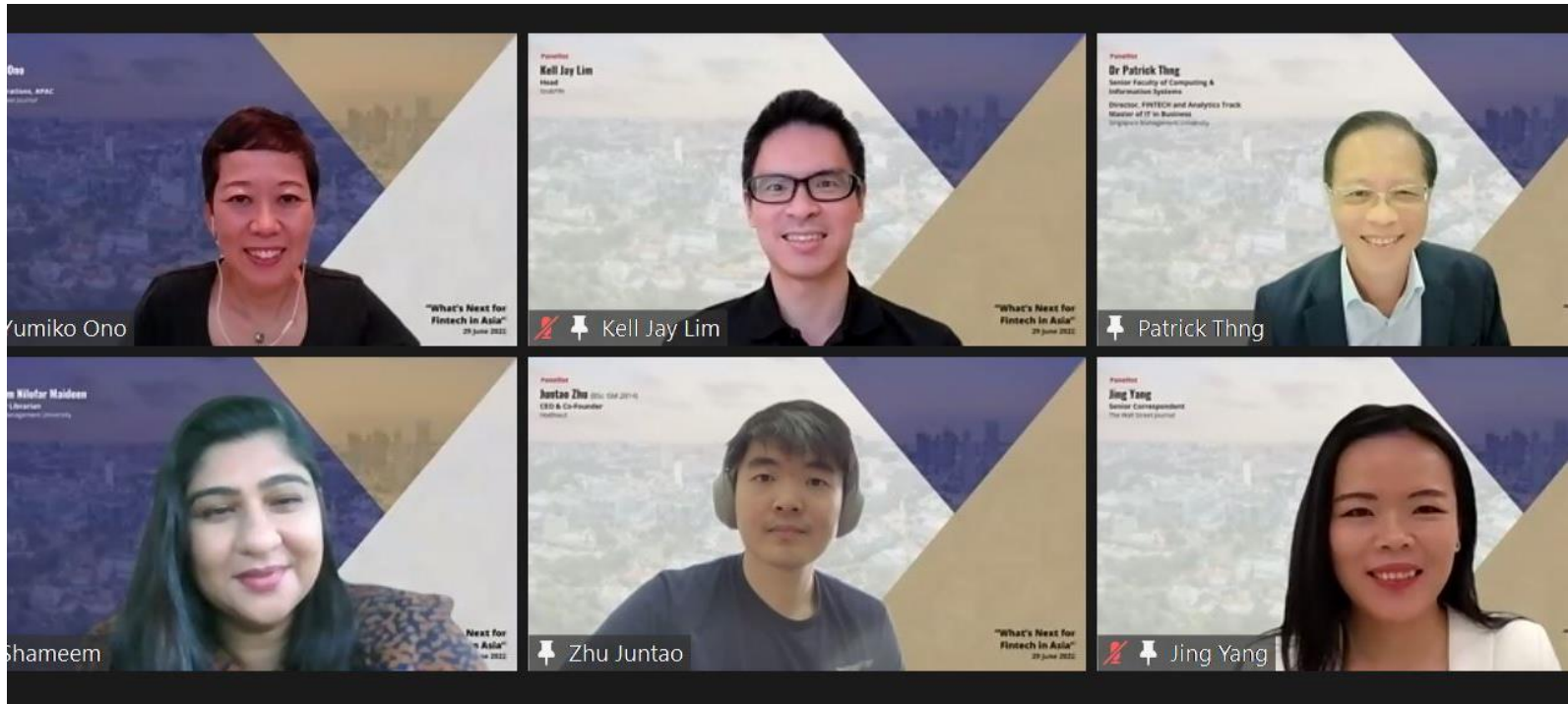
- Discuss implications
- Share best practices
- Engage stakeholders



Other Events and Outreach

Collaboration with schools, alumni, and industry partners to discuss the use of Data Science, AI, and Machine Learning in real world.

- Share industry insights
- Starts conversation
- Connect people



Key Takeaways

It is a relatively new area, so it will take time (and lots of learning!)

If you need to prioritize, prioritize building up the human capital

Engage deeply with faculty/users! They should be our closest allies

SMU SINGAPORE MANAGEMENT UNIVERSITY | Libraries

Thank you!

Questions are welcome

(or contact me at bellar@smu.edu.sg)