

# **BSBINS401 - Analyse and Present Research Information**

## **Session 11: Data Collection and Storage**

Lecturer: Jordan Hill

## Learning Objectives

By the end of this session, you will be able to:

- **Understand methods for accessing and extracting data.**
- **Organize data effectively for analysis.**
- **Implement secure storage practices for research data.**

# Session Overview

- **Topics Covered:**
  - Methods for accessing and extracting data
  - Organizing data for analysis
  - Secure storage practices
- **Activities:**
  - Hands-on data collection
  - Setting up data storage solutions

# 1. Methods for Accessing and Extracting Data

## Data Sources

- **Public Repositories :**
  - [Kaggle](#)
  - [UCI Machine Learning Repository](#)
- **Organizational Databases :**
  - Internal SQL databases
  - NoSQL databases like MongoDB
- **APIs :**
  - [Twitter API](#)
  - [Google APIs](#)

## 2. Organizing Data for Analysis

### Data Structures

- **Dataframes:**

- Library: Pandas
- Example:

```
import pandas as pd  
df = pd.read_csv('data.csv')
```

- **Arrays:**

- Library: NumPy
- Example:

```
import numpy as np
```

## 3. Secure Storage Practices

### Data Security Principles

- **Confidentiality**: Ensure that only authorized personnel can access the data.
- **Integrity**: Protect data from unauthorized alterations.
- **Availability**: Ensure that data is accessible to authorized users when needed.

### Storage Solutions

- **Cloud Storage**:
  - Examples: [Google Drive](#), [AWS S3](#), [Azure Blob Storage](#)
- **Institutional Repositories**:
  - University-provided storage solutions

# Activities

## Lab: Setting Up Data Storage Solutions

### 1. Choose a Storage Platform

- Options: Google Drive, AWS S3, Azure Blob Storage

### 2. Configure Access Controls

- Set up user permissions and access levels

### 3. Upload and Secure Data

- Encrypt data files before uploading

### 4. Implement Backup Strategy

- Schedule regular backups and verify their integrity

## Reading Resources

- **Data Collection Techniques**
  - [Guide to Data Collection](#)
- **Organizing Data for Analysis**
  - *Python for Data Analysis* by Wes McKinney, Chapter 5
- **Secure Data Storage**
  - [Data Security Best Practices](#)



## Lab Resources

- **Tutorials:**
  - [Using APIs with Python](#)
  - [Database Queries with SQL](#)
- **Tools:**
  - [Pandas Documentation](#)
  - [AWS S3 Getting Started Guide](#)

## Next Steps

- **Apply** the data access and extraction methods to your project dataset.
- **Organize** your data following the best practices discussed.
- **Implement** secure storage solutions to safeguard your research data.
- **Prepare** for the next session on advanced data analysis techniques.

## Questions?

- Contact: [jordan.hill@nmtafe.wa.edu.au](mailto:jordan.hill@nmtafe.wa.edu.au)
- Office Hours: Mon–Fri, 9 AM – 5 PM