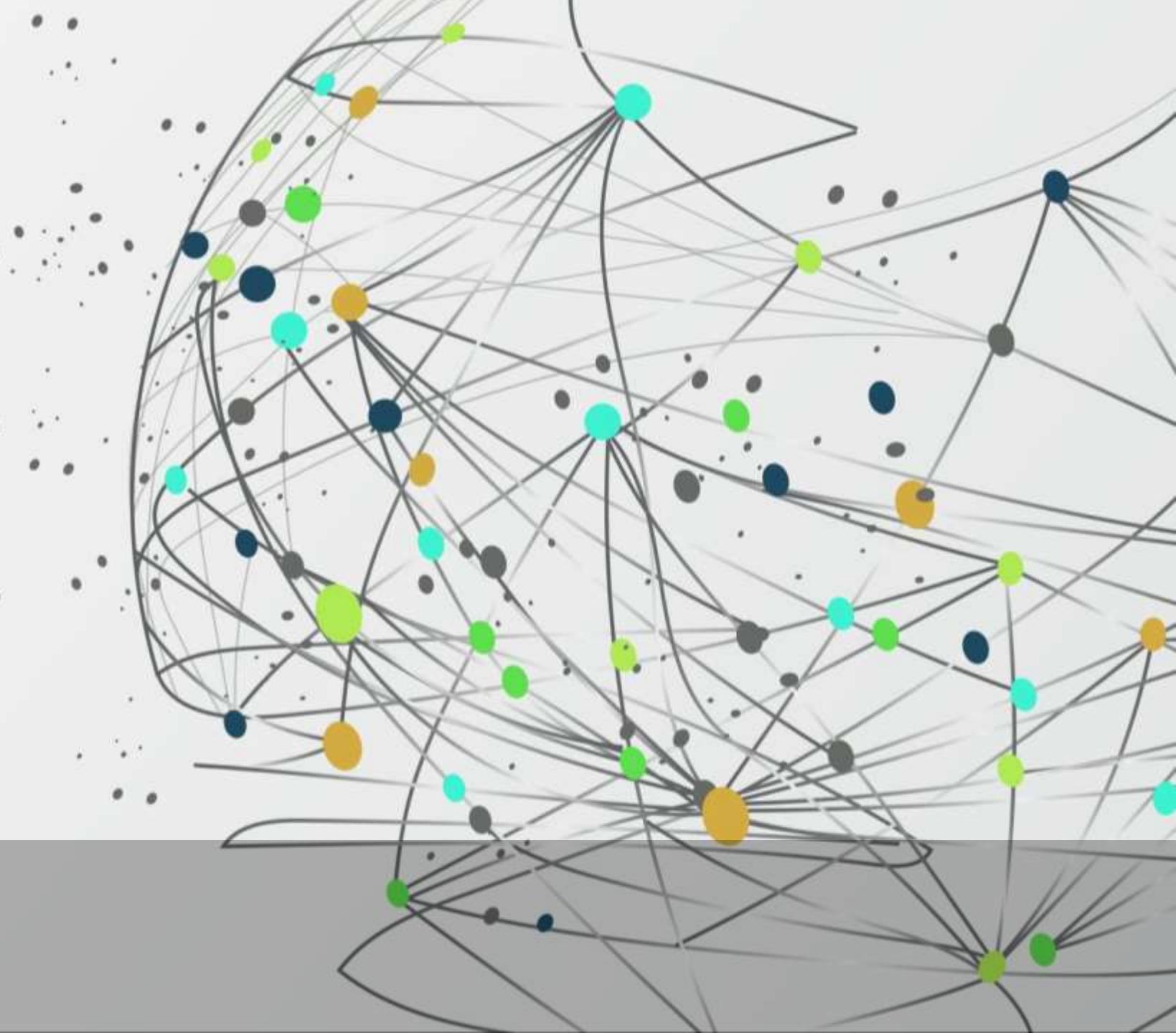


OSINT

Rymer Popovich



Web Scraping

- The process of automatically extracting data from websites and transforming it into structured formats (spreadsheets, tables, databases, etc.)
- Involves sending HTTP request to a web page, parsing the HTML structure to locate the information, and storing that data.
- This technique is very useful for collecting public online data such as prices or statistics when APIs are unavailable.



Web Scraping Example

```
import requests
from bs4 import BeautifulSoup

# URL of the webpage to scrape
url = 'http://example.com'

# Send a GET request to the webpage
response = requests.get(url)

# Parse the HTML content of the page with BeautifulSoup
soup = BeautifulSoup(response.text, 'html.parser')

# Extract and print the text inside the <h1> tag
h1_text = soup.find('h1').text
print('Heading:', h1_text)
```



Metadata Analysis

- The process of examining descriptive information (metadata) associated with datasets/digital files, to uncover hidden patterns or relationships.
- Metadata – includes details about creation date, author, file type, etc.
- By analyzing this data, organizations can improve data governance, detect anomalies, and ensure compliance

Metadata Analysis Example

```
import os
from datetime import datetime

# Path to the file to analyze
file_path = 'example.txt'

# Get metadata of the file
file_stats = os.stat(file_path)

# Extract key metadata attributes
file_size = file_stats.st_size # in bytes
creation_time = datetime.fromtimestamp(file_stats.st_ctime)
modification_time = datetime.fromtimestamp(file_stats.st_mtime)

# Print the metadata information
print(f"File: {file_path}")
print(f"Size: {file_size} bytes")
print(f"Created on: {creation_time}")
print(f"Last modified: {modification_time}")
```

References

- **Metadata Analysis** - *A complete guide to metadata analysis.* (2021, February 16). Canto. <https://www.canto.com/blog/metadata-analysis/>
- **OSINT** - Nordine, J. (2019). *OSINT Framework*. Osintframework.com. <https://osintframework.com/>
- **Web Scraping** - Osborne, J. (2024, April 7). *Intro to Web Scraping: Build Your First Scraper in 5 Minutes*. Medium. <https://medium.com/@joerosborne/intro-to-web-scraping-build-your-first-scraper-in-5-minutes-1c36b5c4b110>