Role of Big Data on Academic Publishing

6th International Conference on 'Scientific Communication in the Digital Age'

Gizem Erboz, Research Assistant

Department of Business and Management, Szent Istvan University

Date: 29.03.2018

What is Big Data?

- Data sets whose size or type is beyond the ability of traditional relational databases to capture, manage, and process the data with low-latency. (IBM, 2018)

- Data could be as

  structured,

  semi-structured

  unstructured data

Trends on Big Data

Source: Google Trends, 2017

Characteristics of Big Data

THE FIVE V'S OF BIG DATA

VOLUME
Vast amounts of data are being generated — the volumes are too large to be analysed with traditional database tools.

VELOCITY
Big data technology can analyze data while it is being generated (at ever increasing speeds) without putting it into databases.

VARIETY
Data is unstructured and diverse — it can be text, images, video, audio, sensors, location data etc...

BIG DATA

VERACITY
Data has different levels of trustworthiness, but the volume makes up for the lack of accuracy.

VALUE
The ability to turn the data into value (insight and intelligence) needs to be included.

Source: CRUK Research, 2017

Figures on Publishing Industry, Big Data or Not?

- >3MM articles submitted
- 300,000 peer reviewers
- >1.5MM articles published
- 30MM readers
- 2B digital article downloads
- 30MM article citations

Big Data for Online Publishers

Collect

multiple repositories, making search quick and simple for online users, whether they're on mobile or desktops
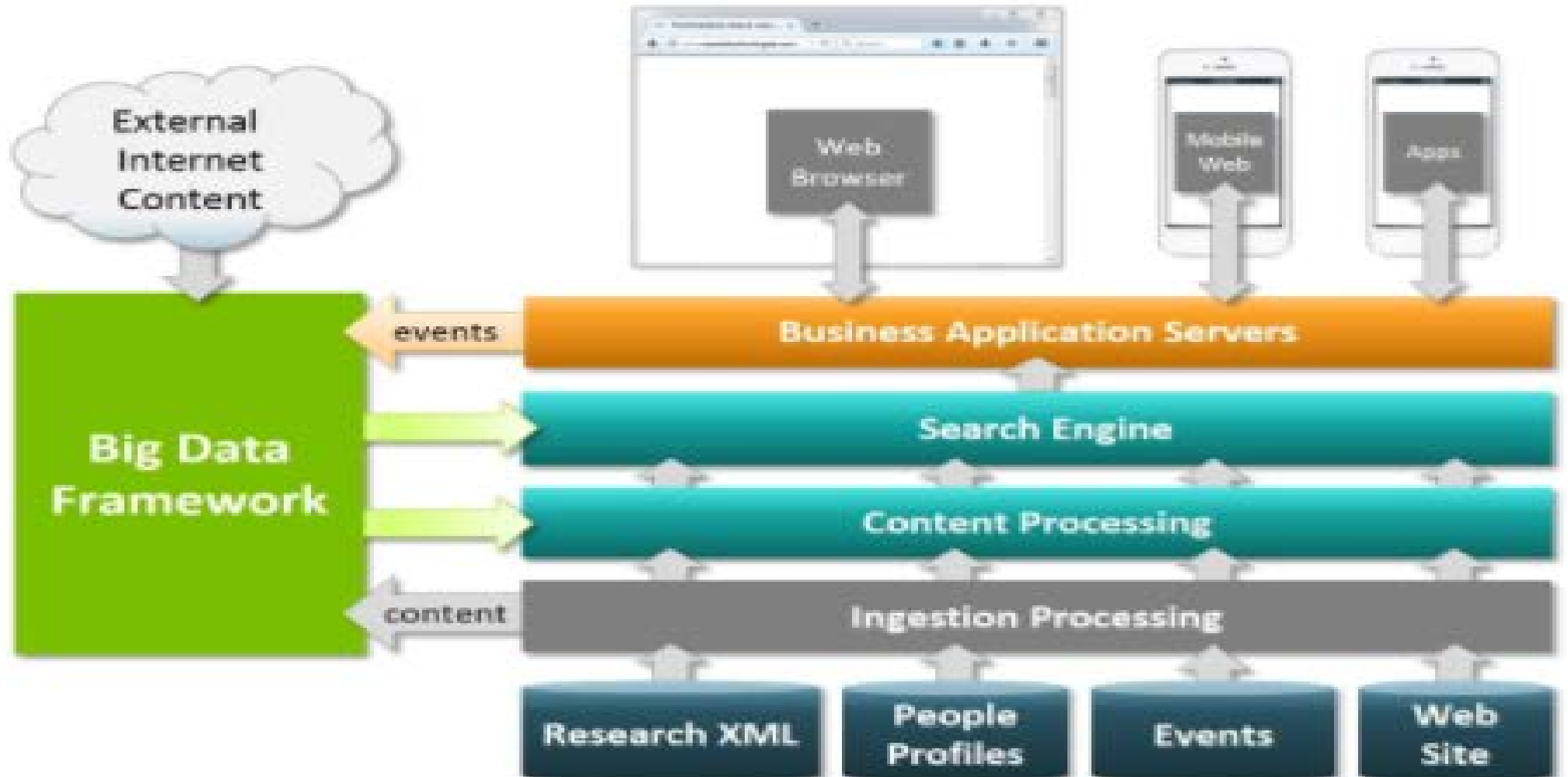
Analyze

Systematically analyze search logs and on-page behavior to "learn" user preferences and personalize content.

Optimize

user experience and sales revenue through search engine testing and scoring

Personalized Content with Log Analytics and Machine Learning

Big data architecture for content personalization

Traditional vs. Online Publishing: Elsevier Case

TRADITIONAL PUBLISHING

- Linear way of Publishing
- Traditional print peer review journals; then were shipped out to academic libraries

- **Problem: Researchers share too much hours for searching**

ONLINE PUBLISHING

- Not just text based documents
- Also, includes content-rich assets such as social media streams, images, audio and video files as well as crowdsourced data
- Larger and Smaller Format Files

- **Problem:valuable data to be processed from relevant sources**

Big Data and Cloud Computing

- High responsibility of Publishers; develop better, faster, more efficient tools and solutions.

- Right content, to right audience, in the right context, when and how they want it.

Elsevier uses Big Data by increasing pools of unstructured data and researchers can easily reach the recommended reading lists and data banks.

'Matchmaker' System

- Predictive Analytics: giving customers what they want, recommended and promoted relevant research

- Other areas where Elsevier employs the applications of Big Data;

- **1. Enhanced content: Article of the Future** which provides a dynamic and interactive reading experience by incorporating tagged and searchable audio files, videos, interactive images and figures, embedded maps, downloadable tables as well as sharing capabilities.

- **2. Re-use of content: TDM, API** to explore patterns across large content databases and derive meaningful analysis from these correlations.

- **3. Solutions content:** delivering not just information and also answers, evidence based insights from massive datasets.

- **4. Sharing content:** cloud based research management and social networking platform, Mendeley, highly increases collobration
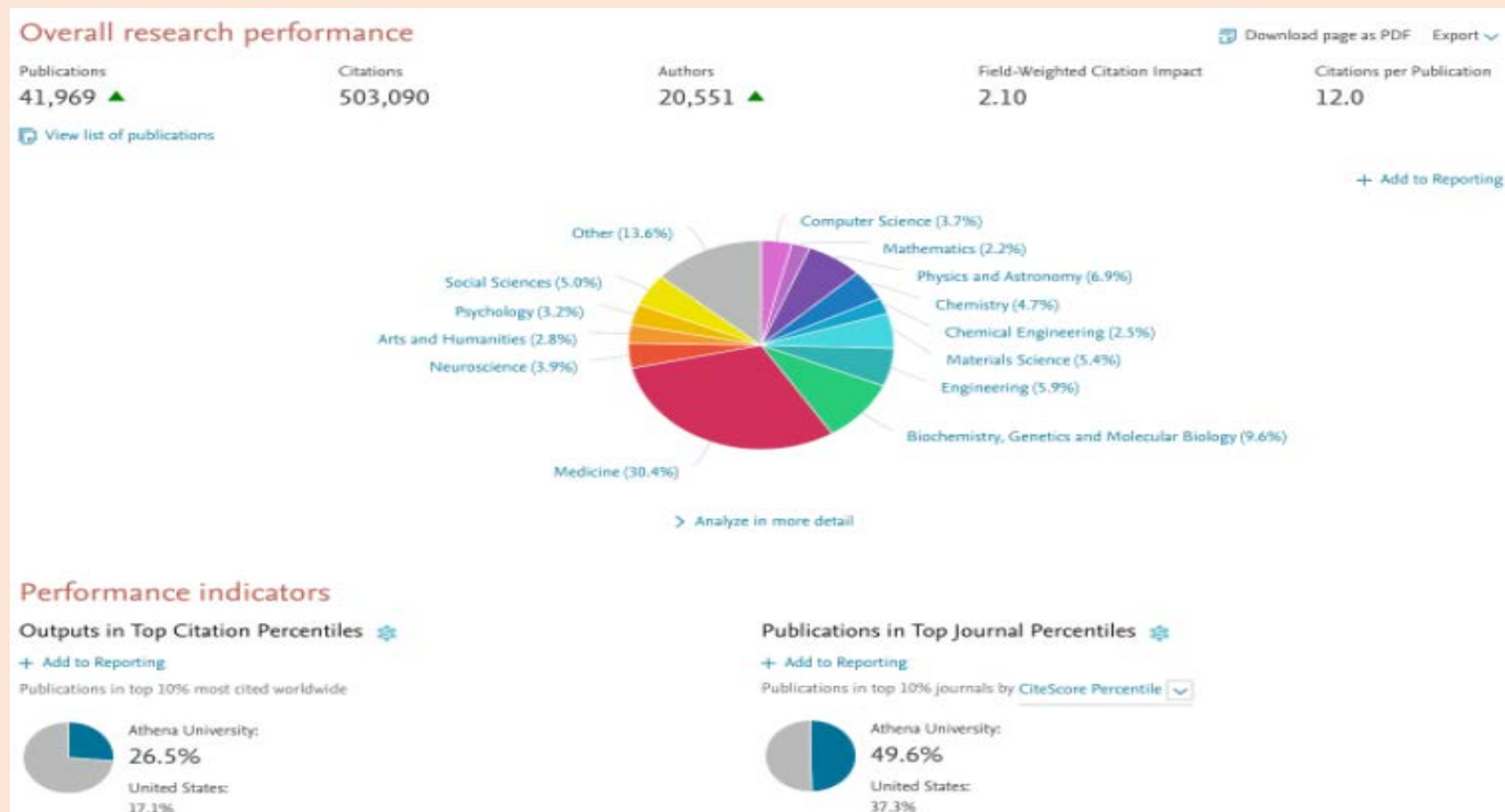
Scival- Elsevier

- easy access to the research performance of 7,500 research institutions and 220 countries worldwide.

- ensure satisfying outcomes, and devise better ways to measure the quality and impact of research projects.

- integrated modular platform that allows you to view and compare any desired research entities such as institutions, countries, self-efined research topics and researcher groups.

- 38 million publication records from 21,915 journals of 5,000 publishers worldwide

- Access to more than 170 trillion metric values to generate powerful data visualizations on-demand, in seconds.
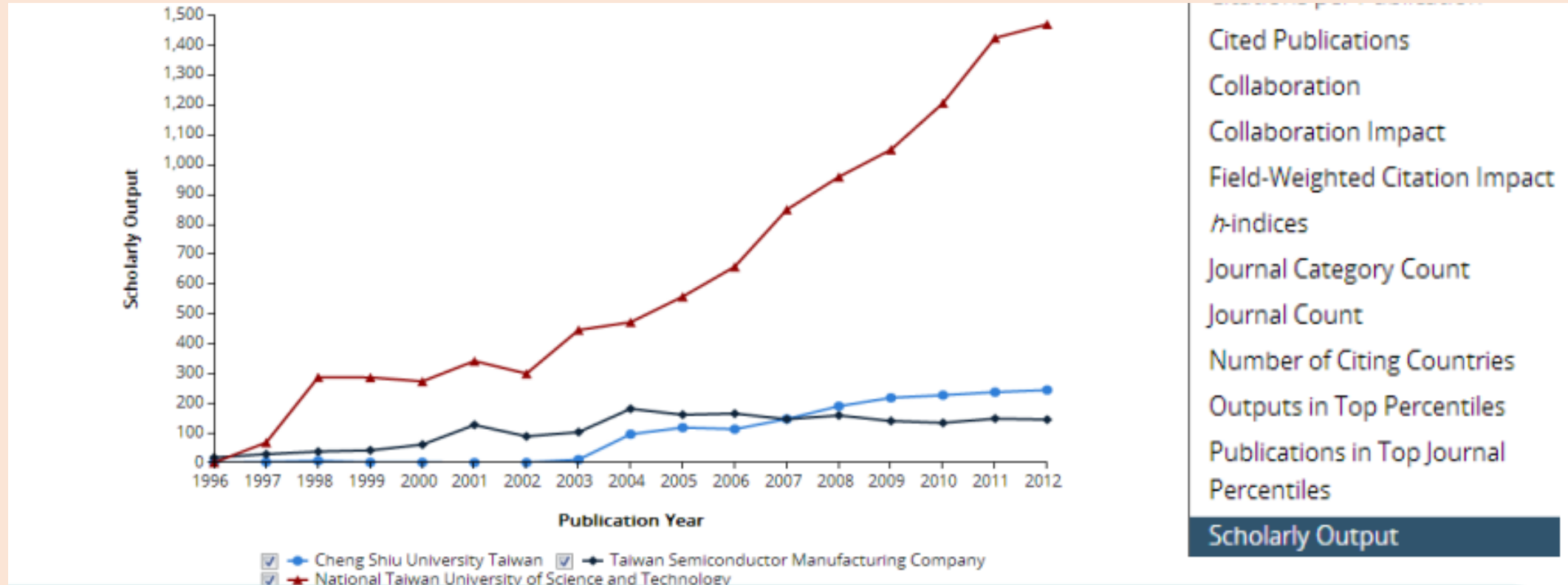
What are the benefits?

- Visualize Research Performance

- Benchmark your progress

- Develop collobrative partnership

- Analyze research trends
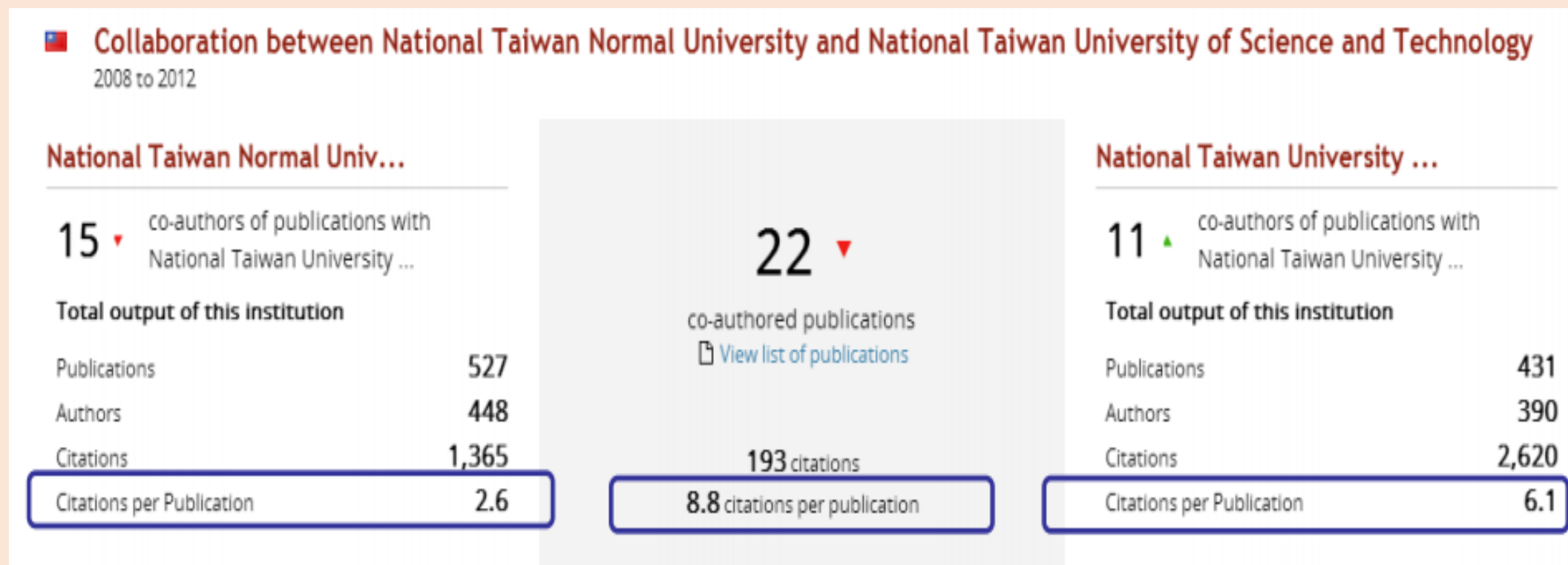
Research Performance

Source: Scival, 2018

Benchmarking: Direct Comparisons

Source: Scival, 2018

Collobration

Source: Scival, 2018

Analyze Trends

Source: Scival, 2018

Thank you