Case Study

Generating a self rehydrating data lake for the educational services industry.

DATA SOCLE

The Smartest way to feed your information systems with sensible data.

Data Socle is business logic driven data processing system that automates extraction, sanitization, customization and usage of data.

The Problem

Data generation & collection in the educational services industry is eclipsed by noise. More often than not, creating a data lake has been an arduous task. Some of the main challenges are :
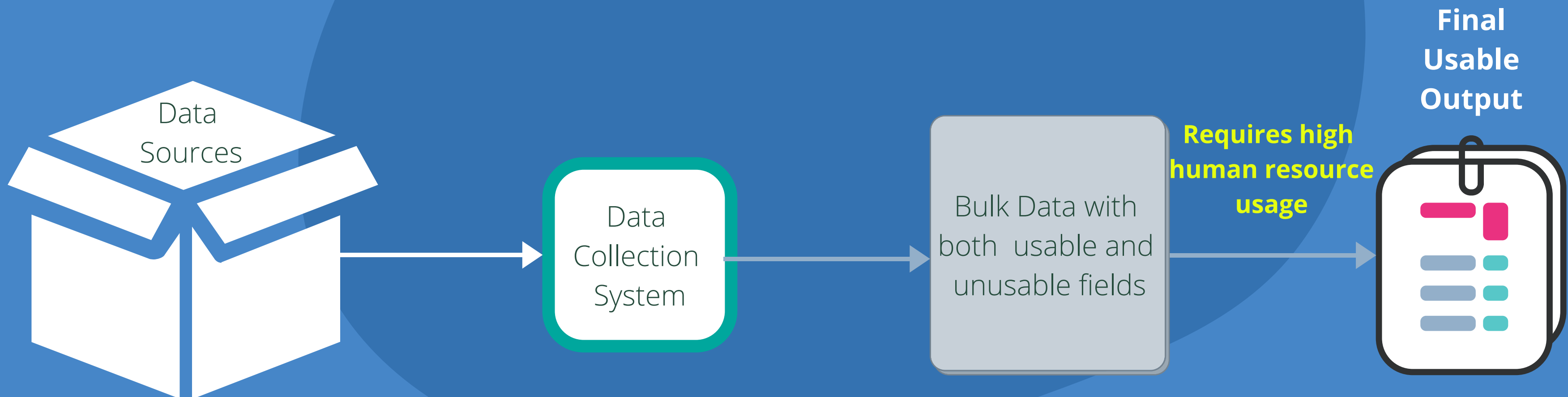
1. Collection of a large percentage of unnecessary redundant data.
2. Inability analyze data due to lack of granularity.
3. Expiration of data and changes in data sources causing data to be unusable.

Redundant Data

Data Collection in the education industry is marred with lots of unusable data points;

For example student entry requirements can be IELTS + SAT for one university while IELTS / SAT for another university. Traditional Data collection systems cannot differentiate between the two.

It takes hundreds of man hours to then sort this data manually and filter out the useful information.

Data Sources → Data Collection System → Bulk Data with both usable and unusable fields → **Requires high human resource usage** → **Final Usable Output**

Lack of Granularity

Some of the most important pieces of data required by organizations like eligibility requirements for scholarships or updates about success stories from alumni cannot be captured by traditional systems due to a high degree of variance in the format, design and presentation of the data source (The University website or social media handles)

Losing out on vital information like this can eventually lead to a drop in student engagement rates.

Static collection systems that do not use pattern recognition or machine learning

Data Sources

Data Collection System

Collects only High Level Data

Final Output missing key elements
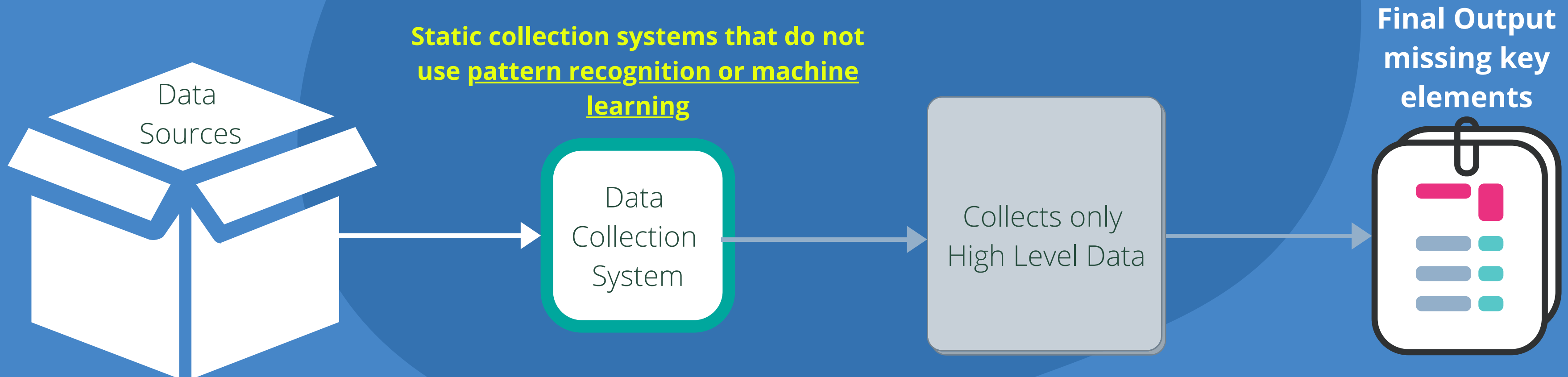
Expired Data & Sources

Perhaps the most persistent problem that consumes the largest amount of time and capital in the educational services industry is tied to it's inability to keep their data lake updated. As Universities update their websites, course details, fees and various other components of their educational offering, some of the biggest players in this sector provide data that is atleast 30+ months old.

Traditional Data Collection Systems cannot detect changes and updates in the data source, which means expired data remains in the system for a long time.

Data Sources

Cannot Detect updates and changes in Data Source

Data Collection System

One time collection of bulk data without data update protocol

Expired Data
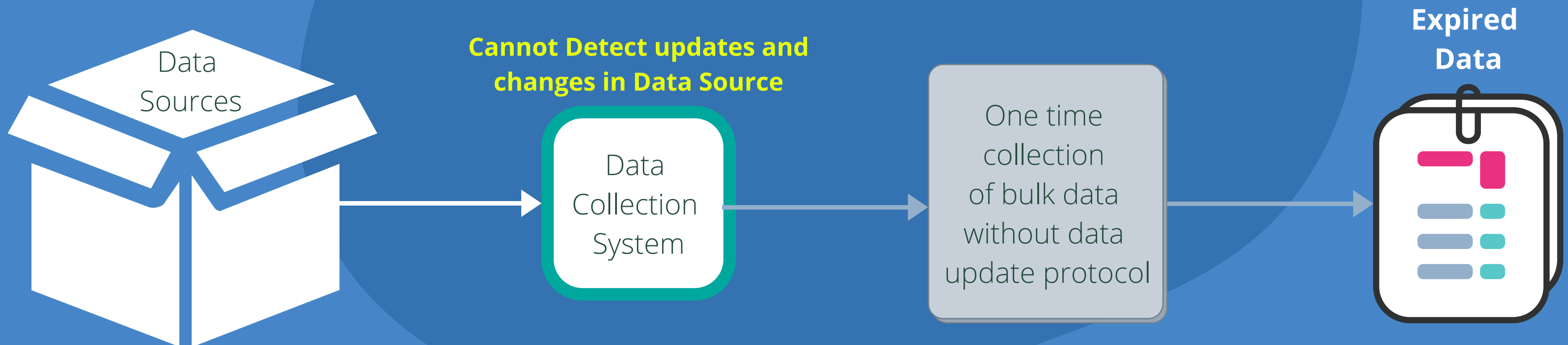
We've used Data Socle to solve these issues.

By using Data Socle's Artificial Intelligence core, powered by Machine Learning and Pattern Recognition, we have built a system that automatically :

1. Recognizes & extracts the required, useful data from a preplanned template. (Zero Redundancy)
2. Uses pattern recognition to transcend the barriers of formats and localization to provide the most granular data, quickly and consistently;
3. Checks the data sources frequently to ensure all data points are validated and updated.

Results

Automation

The customer can now use Data Socle to directly extract, sanitize, customize and reuse data about universities in their website. We have automated this flow such that, there is no requirement for any human to validate or check the data. Data Socle automatically refines the data and provides it in a JSON format used in the website.

Updation

We've integrated Data Socle to the customer website such that, any changes detected in the data sources is automatically updated in the website on a weekly basis. There is no question of expired data and we at Data Socle can ensure data relevance.

Analytics

By providing a set of highly granular error free data, our customer is now able to reliably use the data set to run advanced analytics which was impossible with the previous system. Data Socle has enabled our customer to provide more value and engagement to students by providing unique insights and enhanced information.

The Financial Perspective

A quick summary of the amount of capital our customer in the education industry was able to save through this initiative with Data Socle.

₹ 2.8 Million

In terms of Data Collection and Entry professionals

₹ 1.4 Million

In IT Services to build, update and maintain Web Scrapers

1800+ Man Hours

In labor, project management etc.

How does this work?

Unlike traditional data collection systems, we at Data Socle are not in the business of just collecting and providing data, but we go a large step further in ensuring that the data is updated, usable and constantly available.
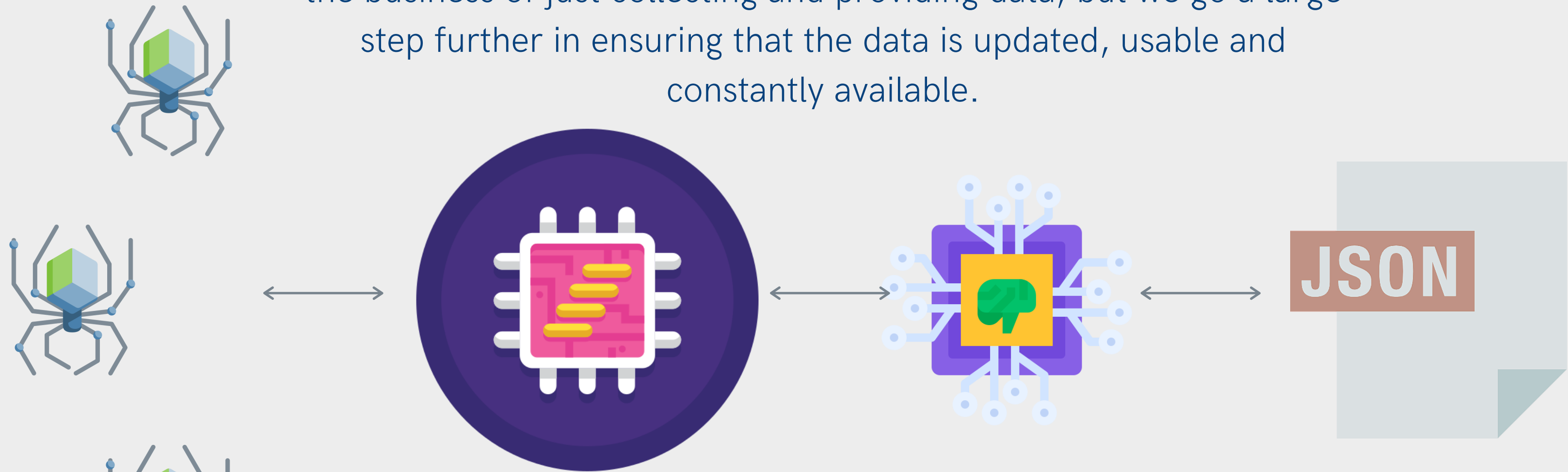
Our crawling spiders powered by a proprietary Machine learning core analyzed and crawled 1000's of data sources, from more than 1600 universities to detect and formulate a pattern to collect the most granular level of data.

The Data Processor then filtered the data into a structured format (JSON) in this case, with all the necessary information being automatically updated on a monthly basis.

Traditional Data Collection Systems

DATA SOCLE

- Requires constant human intervention and effort.
- Requires constant updates and maintenance.
- Delivers a result that requires rework.

- One time setup, requires little to no maintenance.
- It's fully automated, updated data is guaranteed.
- Integrates with any existing information system seamlessly.

Education Industry & Beyond

Data reserves is as important as financial reserves in 2020. In this day of data running business decisions, building a data lake is the utmost priority of CXO's across the world.

If you are in the Education / Educational services industry looking to provide more value to your business or to your customers through data science, Data Socle is the smartest solution available in the market today for you to build a data lake and keep it hydrated.

Setup a demo with us or log on to https://datasocle.com/ for a free demo. We take pride in the fact that we can help you look at data with a fresh, new perspective.

Sample Data Set

```
[
  - {
        University Name: "MCMASTER UNIVERSITY",
      - University Rank: [
            "144"
        ],
        Type: "Public",
        Accreditation: null,
        Scholarships: "Yes",
        Country: "Canada",
        Region: "North America",
        Website: https://www.mcmaster.com/,
      - Google Maps: [
            "https://www.google.com/maps/search/?api=1&query=Main campus 1280 Main Street WestHamilton, Ontario L8S 4L9Canada"
        ],
        Student Count: "25,008",
        Faculty Count: null,
      - Address: [
            "Main campus 1280 Main Street WestHamilton, Ontario L8S 4L9Canada"
        ],
        Description: "McMaster University (McMaster or Mac) is a public research university in Hamilton, Ontario, Canada. The main McMaster campus is on 121 hectares (300 acres) of
        land near the residential neighbourhoods of Ainslie Wood and Westdale, adjacent to the Royal Botanical Gardens. It operates six academic faculties: the DeGroote School of
        Business, Engineering, Health Sciences, Humanities, Social Science, and Science. It is a member of the U15, a group of research-intensive universities in Canada.The university
        bears the name of William McMaster, a prominent Canadian senator and banker who bequeathed C$900,000 to its founding. It was incorporated under the terms of an act of the
        Legislative Assembly of Ontario in 1887, merging the Toronto Baptist College with Woodstock College. It opened in Toronto in 1890. Inadequate facilities and the gift of land in
        Hamilton prompted its relocation in 1930. The Baptist Convention of Ontario and Quebec controlled the university until it became a privately chartered, publicly funded non-
        denominational institution in 1957.\nMcMaster University has over 27,000 undergraduate and over 4,000 post-graduate students. Alumni and former students reside across Canada
        and in 139 countries. Its athletic teams are known as the Marauders, and are members of U Sports. Notable alumni include government officials, academics, business leaders,
        Rhodes Scholars, Gates Cambridge Scholars, and Nobel laureates.",
        Email: null,
      - Phone Number: [
            "+1 905-525-9140"
        ],
        Logo:
        "/9j/4AAQSkZJRgABAQEBLAEsAAD/2wBDAAUDBAQEAwUEBAQFBQUGBwwIBwcHBw8LCwkMEQ8SEhEP\nERETFhwXExQaFRERGCEYGh0dHx8fExciJCIeJBweHx7/2wBDAQUFBQcGBw4ICA4eFBEUHh4eHh4e\nHh4eHh4eHh4eHh4...
      - Map Coordinates: [
            "43.26 Deg N",
            "79.92 Deg E"
        ],
      - Courses: {
```

Sample Data Set

```
                "+1 905-525-9140"
        ],
        Logo:
        "/9j/4AAQSkZJRgABAQEBLAEsAAD/2wBDAAUDBAQEAwUEBAQFBQUGBwwIBwcHBw8LCwkMEQ8SEhEP\nERETFhwXExEaQaFRERGCEYGh0dHx8fExciJCIeJBweHx7/2wBDAQUFBQcGBw4ICA4eFBEUHh4eHh4e\nHh4eHh4eHh4eHh4
- Map Coordinates: [
                "43.26 Deg N",
                "79.92 Deg E"
        ],
- Courses: {
        - UG: [
                - {
                        Course Name: "Arts & Science",
                        Eligibility / Entry requirements: "10+2 complete and IELTS, TOEFL, CAE, PTE, CAEL, TOEFL iBT",
                        Course Intake: null,
                        Course Fee Canadian/Permanent Residence: "CAD 201 per unit",
                        Course Fee International: "CAD 1,079 per unit",
                        Course Duration: "3 years",
                        Application Date Summer Term: "April to May",
                        Application Date Winter Term: "Jan to Feb",
                        Application Fee: "CAD 110 - CAD 150"
                },
                - {
                        Course Name: "Bachelor of Technology",
                        Eligibility / Entry requirements: "10+2 complete and IELTS, TOEFL, CAE, PTE, CAEL, TOEFL iBT",
                        Course Intake: null,
                        Course Fee Canadian/Permanent Residence: "CAD 201 per unit",
                        Course Fee International: "CAD 1,079 per unit",
                        Course Duration: "3 years",
                        Application Date Summer Term: "April to May",
                        Application Date Winter Term: "Jan to Feb",
                        Application Fee: "CAD 110 - CAD 150"
                },
                - {
                        Course Name: "Business",
                        Eligibility / Entry requirements: "10+2 complete and IELTS, TOEFL, CAE, PTE, CAEL, TOEFL iBT",
                        Course Intake: null,
                        Course Fee Canadian/Permanent Residence: "CAD 201 per unit",
                        Course Fee International: "CAD 1,079 per unit",
                        Course Duration: "3 years",
                        Application Date Summer Term: "April to May",
```

Contact Us for a Demo, Proof of Concept or Custom Requirements.

+91 8880171000

sales@datasocle.com

https://datasocle.com