## ***Framework for Building Web Scrapers***

### ***By: Paul Paskevicius***

Scrapers are code scripts, primarily written in Python, that scrape specific data from websites, blockchains, protocols, PDFs, software, etc. While scrapers are very beneficial, they are complex and time consuming to build. **This framework will help determine whether a certain project should use a scraper or not.**

---

What, where, when, and how?

This should be done first to help define the scope of the project:

- What data is being scraped?
- From where is it being scraped and where is the data going?
- When will the project be complete and how often is the scraping happening?
- How is the scraper built? What language, what system or PC?

Value or the Why?

The most important question to ask is why? Why build a scraper:

- Is the scraper going to help us learn something new which will lead to a more efficient outcome?
- Is the scraper going to generate alpha? If yes, how is it going to generate alpha? Detail from beginning to end how the scraper will ultimately result in alpha.
- If not alpha, is it going to automate something? Will it allow us to build, improve or define better parameters of something? Will it help us create a system for how we interpret something, like the markets or tokenomics, etc.?
- Will it make current systems more efficient? How much more efficient?

Time

If any of the above had clear yes answers, then that leads us to the final factor: time. We have to determine how much time it would take to build the scraper vs how much time it would save in the long run.

- Estimate the amount of time it would take to build this scraper, include everything from learning the code to final implementation of the code: _____
- Estimate the time that would be saved in the future: _____
- If time saved > build time, then the scraper should be built and implemented
- The final caveat is that although most scrapers will be designed to save time (since we are automating and increasing efficiency), there could be rare instances where a scraper will not save time but it still needs to be implemented. If this is ever the case, then the scraper should be built