The move towards data-driven science has resulted in a data explosion

Scientific file systems and data repositories continue to grow at alarming rate

Repositories quickly become data swamps if without considerable continued investment

Poor organization and clutter in large scientific repositories complicates data discovery and analysis

We can see from these example. We have data in different formats, different extensions, different repositories.

It’s hard to find desired data in all of these files. The different extensions may become annoying, you may get lost in all of these directories. How can you know where is your data among all of the similar files?

To mitigate the effects of high-velocity data expansion and dis-organized data repositories, we develop

Skluma –

An automated metadata extraction system that extracts rich content and context-based metadata from scientific files

Able to extract from embedded structured data, free-text documents

Process extracting metadata and inferring contextual relationships between files in such repositories, converting them into indexed, searchable collections

