

WHAT SCIENCE MAY BE

by **Breck Yunits**

APRIL 3, 2025

We can embody all of science into a single fully connected text file.

Scientists would contribute blocks to this file like they now contribute papers to journals and datasets to databases.

I estimate all of science would fit in 1 billion pages. If printed this would extend from here to the Moon.

This file would allow everyone to get state-of-the-art, logical answers to any question by querying every scientist, living and dead, all at once, on their own machines.

*

How do we connect every word? A new language consisting of a simple universal syntax and list of objects called parsers would create an invisible wired grid.

This file would be topologically ordered, meaning concepts are defined only out of already defined concepts. For example, addition and uncertainty would be defined before quantum mechanics.

All words are typed and the file is mostly structured data, with exceptions for sections of visual and free-text data, called groundings.

Experimental procedures and actual scaled measurements would be found early and often, as reproducible experiments are the bedrock of science.

*

To attract the huge energy required to build this file it must deliver value far before it is near complete. Domain specific sections of the file can be built separately in parallel and deliver immediate value even before integration. The tools needed to build this system are also useful for more down-to-earth tasks and these use cases could attract the energy necessary for the moonshot. For example, the same software needed for this file could also power simpler and more trustworthy blockchains.

Decentralization and forks would be encouraged and often merged back, and may on occasion lead to frac-

turing and different schools of science.

Believers in this system would want Freedom of Science laws that would protect the rights of individuals to create, improve, and share these files.

This system might be built by ten million scientists contributing an average of 100 pages each, or perhaps a far smaller team utilizing AIs.

*

Think of these files as a new medium beyond articles, books, databases, encyclopedias, wikis, and LLMs.

The three major alternatives to this system, in order of connectedness, are libgen, which metaphorically glues all liberated books and scientific papers together at the edges with no integration; Wikipedia, which contains a shallow but broad collection of concepts with weak integration; and deep neural networks, which turn libgen, Wikipedia, and the web into an inscrutable matrix with incredible generation capabilities that strongly suggest embedded logical understanding (but also high energy requirements and hallucination tendencies).

This system differs from previous expert symbolic AI systems in the design of its language that allows it to scale to all of science.

*

This file would instantly reveal what science knows and does not know; what concepts are needed to fully understand other concepts; what experiments one can do to verify key concepts; and what the actual contributions of new research are, measured in line changes.

In its transparency, simplicity and minimalism, this file would make science a physical thing that people could hold and trust.

A number of developments have made it more feasible than ever to build this system now, but the value this system would provide to humans is timeless.

**