

The Amnesiac Civilization

David S. H. Rosenthal
Stanford University Libraries

What is the data that informs a society? It is easy to think that it is just numbers, timely statistical information of the kind that drives Google Maps' real-time traffic display. But the rise of text-mining and machine learning means that we must cast our net much wider. Historic and textual data is equally important. It forms the knowledge base on which civilization operates.

For nearly a thousand years this knowledge base has been stored on paper, an affordable, durable, write-once and somewhat tamper-evident medium. For more than five hundred years it has been practical to print on paper, making Lots Of Copies to Keep Stuff Safe. LOCKSS is the name of the program at the Stanford Libraries that Vicky Reich and I started in 1998. We took a distributed approach; providing libraries with tools they could use to preserve knowledge in the Web world. They could work the way they were used to doing in the paper world, by collecting copies of published works, making them available to readers, and cooperating via inter-library loan. Two years earlier, Brewster Kahle had founded the Internet Archive, taking a centralized approach to the same problem.

Why are these programs needed? What have we learned in the last two decades about their effectiveness? How does the evolution of Web technologies place their future at risk?