Reprinted in Essays of an Information Scientist: Awards of Science, Vol:7, p.346, 1984.

Also see Introduction on p.346, Essays, Vol:7

Information as a Communicable Disease

by Kevin Kelly

The history of science is a house of upsets. Here comes Galileo banging down the doors of astrology, here comes Darwin dragging out the attic furniture, here comes Einstein ripping up curtains to let the light in. To the faraway eye the whole structure is constantly shifting, unfolding, changing its floor plan just when you thought you knew its shape. Even the mathematical basement could overturn. Things don't stay put for more than two minutes in science. There are revolutions brewing in the hallways!

The system makes sense in the end, but who can stand the excitement?

The Truth is not so radical. What actually happens, happens in between the in-betweens. Information growth is so constant and finely textured that it blurs completely when we look back. A hundred years later, only the dramatic gestures are apparent. We can't see the real push and shove of science unless we bend and face it daily from two inches away.

In the research lab where I work, at least, we never see any revolutions, and we've been looking. Things plod along so slowly that, if there's any excitement, it's akin to the beauty of a Japanese play where almost nothing happens. Gloria F. has run the same experiment, with minor variations, nearly a thousand times in the last five months. You can't expect to get your socks blown off with a tame routine like that.

Scientists are driven to do the same old stuff by simple curiosity—the persistent, childlike variety. They are bugged to death when they don't understand something they think they could. The most common refrain I hear in the lab, probably once an hour, is these exact words: "Why does it do this? Why does it go like that? What will happen if we do this?"

The bookkeeping needed to track the answers is regimented and standardized. There is nothing very revolutionary about adding another entry in the ledger. Science is almost uniformly unoriginal because it is incremental by design and derives its value primarily from previous entries. Nothing would be worse than to invent a new number symbol or revise a tallying method in the middle of a book. Things must happen one line at a time.

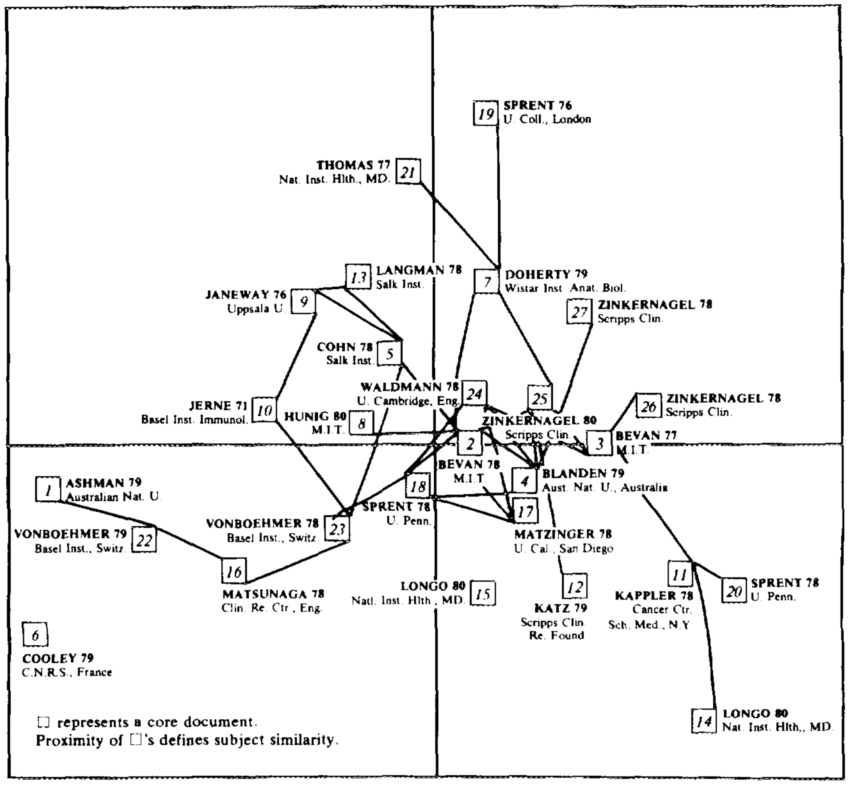
Life in the lab would flatten out into complete boredom without the little mysteries and hard-to-explain findings which always come up. These give us a chance to ask "Why?" and spend time wondering. Most mysteries are eventually explained, if not by us, then by someone in Massachusetts or Florida. But there are always anomalies that don't go away and that we have to work around. If enough of these pile up in one area so that it becomes awkward to work there, then a curious thing happens: some whippersnapper butts in with a new way of accounting which everyone reluctantly agrees will take care of the discrepancies. The data is transferred, the whole show is run from a different book, and the new system is the standard until anomalies pile up again.

With keen vision you can see these changes coming like comets. I was inspired to work for a university by a guy I've never met who codiscovered the structure of DNA and won a Nobel prize for it. He was trained as a bird watcher but saw a change coming and shrewdly switched his line of work to biochemistry, perhaps because he lacked preconceived ideas of what the new perspective had to look like. I gathered from this that any normally brilliant fellow who allowed himself to be sensitive to creative ideas could, by being in the right lab at the right time, luck into some awesomely simple insight. It was just a

In addition to traveling all over the globe and founding Nomadics Books in Athens, Georgia, Kevin Kelly has spent a fair amount of his time in college libraries. He wrote this examination of the flow of scientific information following a year working for the University of Georgia, co-producing a medical education and research film about human digestion. —*Jay Kinney*

Reprinted from: *CoEvolution Q.* 42:98-103, 1984.

Immunological diversity of T-cell-response in radiation-chimeras



A map showing the principal infectors (cited authors) in the epidemic of current research on the immunological diversity of T-cell response in radiation-chimeras. Through their citations, infectees (citing authors) identify the vectors (cited papers) that communicated the germ of research to them. The most concentrated groups of papers reflect the most similar research subjects. (From *ISI Atlas of Science®: Biotechnology and Molecular Genetics, 1981/82* and Bibliographic Update for 1983/84.)

matter of hanging loose. If you didn't harden your mind and you kept on your toes—ready, like an ace short-stop—you could grab the Big One as it zoomed down from the heavens, and score the gift, just because you were ready! I woke up after six months in the lab and realized that I had as much chance of uncovering some Nobel truth as I did of inadvertently building a finely crafted cabinet or accidentally learning to play the piano. The grain of discovery doesn't run that way.

That there was a texture, a geography, a structure to what we knew, how we knew what to look for and the rules to follow to let us know when we found it was news to me.

Some of the lumpiness in what we know is caused by the special way scientific information is passed along. If every time I passed an idea to you I was required to mention where I first got a hint of it, we'd have a kind of scientific notation. When you passed on the idea I gave you, modified by a few others you picked up from someone else, you would mention that you got some ideas from me. If we did this in writing, and kept all of it, you wouldn't have to mention who gave me my idea, because anyone could find my letter to you and look it up. If we were very orthodox we would do this for every idea, every time. We would soon have whole libraries full of stuff that few people would want to read.

And that's what our libraries have today, whole shelves of books that are never read. Strangely, the better the library, the more complete their records, the more books they have that nobody reads. In a book that hadn't been checked out of the university library in three years I read a study that revealed that only half of the books in a library are checked out in a given year. Another rarely referred-to journal stated that less than half the back journals are referred to each year. And a little-noted paper demonstrated that half of the two million papers published each year are never noted by anyone. The statistics would imply that a few papers might not even be read at all. Yet papers that are never read are kept because their mere existence is valuable. A whole chain of better-read papers is built upon them the way important sums depend on petty figures.

The result is interesting, because papers and books are the presumed currency of information exchange. Most papers are ignored because of the old 20-80 effect, or in this case, more like the 10-90 effect: ten percent of the people do 90 percent of the work. An investigation of documents concerning milkweeds claimed that all the information to be found on the subject was contained in 96 of the four thousand or so papers written on the subject. Until recently the bugaboo was that the researcher wouldn't know which 96 until he finished reading all four thousand.

Scientists won't dump the 39 repetitious papers found in every 40 because redundancy is the sister of accuracy. The price of precision is a cumbersome mess of numbers and papers. So what, the scientist thinks, you just archive it in libraries, which they tend to view as bottomless filing cabinets. Librarians know better. They know they couldn't possibly keep all two million papers published each year.

Books, once ordered, take a long time to arrive, and to be most useful should be in the library shortly after being printed. A couple of librarians will be responsible for securing all the best, newest, most durable information available to science. This means librarians are peeking at the edges, all edges, of science, trying to discern which information is most useful in areas they couldn't possibly be expert in, and to have it ready when we need it.

So many books flow though librarians' hands that they pick up the fabric of what's happening. They know what's hot, what's stagnating ("We're quite rosy in biomass these days"). They learn the rough geography of the expanding terrain from what comes in, and wherever the map is empty they start filling it.

Librarians used to purposely leave blank spots to save money, then borrow information from another library when necessary. These days the real cost of interlibrary loans is so expensive that it is cheaper to buy the book. Our library will automatically purchase any book I request if it costs $100 or less. Science books published last year averaged $51 apiece (some go for $200). That works out to 11 cents per page, which makes the five-cent copy machines in the library an incredible deal.

The two million scientific papers published each year appear in about 80,000 different sources. My friends in the labs each read about 200 of these articles per year. The main method they use to connect with the 200 papers that will do them any good, while weeding out the thousands of redundant ones and the other million or so that have nothing to do with them, and at the same time not missing important ones that slide in at different angles, is to read a tiny, unusual magazine called *Current Contents*®.

This ingenious magazine is compact, about the size of a *TV Guide*. The whole thing is nothing more than the reproduced tables of contents from the several thousand best scientific journals. Table of contents pages in the journals normally list titles and authors of formal, rigorously styled reports. In a single issue of *Current Contents*, about 3,000 papers will be listed. There are seven flavors: life, clinical, social, physical, environmental, engineering, and the arts. The professor I work with eagerly awaits the arrival of *Current Contents* and steals them home to read in bed at night, going through about 1,000 titles each week.

According to a paper recently listed in *Current Contents*, "For every person who reads the whole of a text of a scientific paper, twenty read through the summary, and 500 read the title and stop there. Most papers have their title read and no more." Therefore, a lot must be conveyed by the title. Tremendous energy is spent squeezing the whole message of the study into a brief and appealing headline. In fact, the sensation of reading *Current Contents* (once you have the lingo down) would be similar to reading a weekly Sunday magazine that was composed solely of headlines of the major stories from newspapers around the country. (Not a bad idea for a magazine actually.) Some headlines would be more informative than others, but on the whole you'd get a funky sense of what was going on. You could read the full story by looking up the proper day's paper.

As scientific titles become less of an indication of the paper's content and more of an advertisement, the temptation is toward yellow journalism. In the same vein that supermarket newspapers claim "Study Shows There's Life After Death" or "Lose Weight Eating," so scientists are tempted to work into a title catchy hooks like *cancer* or *cloning* or other best-selling buzz words that are only indirectly connected to the main point of the article. The art of title writing is so important that the title is often the most revised portion of a paper. On occasion, workshops devoted to honing this skill are offered at universities. Some authors completely submit to a pragmatic view and figure that since most folks are going to get no farther than reading the title, the title should give, as much as possible, the results or conclusion of the study. Instead of "A Preliminary Theoretical Model of User Patterns in Scientific Information," just sock it to them: "Texture of Science Is Like Lumpy Gravy."

The author's address is indexed in the back of each *Current Contents*. A researcher baited by the title of a paper mails a post card to the author, requesting a reprint, which the author sends at no charge as long as the supply lasts, which is often not long if the title promises a hot one. The author would like to have an unlimited supply of reprints but he must pay for them himself, as well as often pay to have his article published in the first place. (This is another story.)

Requested reprints trickle back to the research labs, are scanned, and occasionally actually read, and then are filed away in microlibraries according to each scientist's idiosyncratic system. My friend's basic research library is two large filing cabinets that hold all the world knows about pancreatic lipase enzymes. They are within arms' reach sitting down. Half are free reprints, the other half are Xeroxed from back copies of journals in the library.

These latter are acquired by going on a once-a-week or once-a-month library run. Clutching an ongoing grocery list of titles, our hero dashes off to the main library, scurries through the stacks collecting journals, hurries to the copy machines, reproduces them, grabs the copies and runs back to the lab. It is as if scientists didn't like being in libraries. They never read anything there, at least I've never caught one, and it's a rare sight to see one browsing, even for pleasure. Libraries are for students who don't know any better. Yet researchers love to be near libraries. The principal reason why the majority are working at institutions instead of being wired up to an electronic cottage among the cows is because they can run through the big libraries any time they need to.

Some third-world researchers depend on *Current Contents* and the free reprint system for up to 90 percent of their information base. In some disciplines they are the main users of the system. There's a box in the lab overflowing with stamps from around the world. The most exquisite ones were pasted on post cards with polite requests for the latest on "The Hydrolysis of Triacylglycerol Emulsions by Lingual Lipase," a reprint from our lab, to be sent to India or some other distant place.

The future is computers, but they haven't arrived in the science labs, yet.

Certainly the bulk of laboratory apparatus is programmed with computer chips—the scientist couldn't do a thing without them now—but people in the labs are almost unanimously skeptical of the computer's role for their information. They claim that less than 1 percent of their useful stuff is learned by computer networks. The infrequent computer searching they do is chiefly to double-check what they already suspected.

Telephoning is big, however. John P. will be struggling through a poorly written paper, throw up his hands in frustration, and call up the author. How in the world did you come to that conclusion? he demands. Nearly every scientist I've spoken with has praised the casual table-talks following a conference as their most important source of new ideas. That is when the meaty news comes, unburdened by sticky references or cautious statements. Award-winning biology-watcher Lewis Thomas observed, "I have the impression that a great body of information is getting around by a mechanism that can only be termed gossip."

Maybe we're in the middle of a gossip explosion. The facts show that the paper-per-author output has remained steady, implying there is not a literature explosion so much as a population (of scientists) explosion. More scientists, more hubbub. At any rate, there are more papers and more new ideas flying around than before, and it has taken one or two of these new ideas to tame the rest and make them accessible. Let me try to explain one of them.

A scientist studies to prepare his brain for a good idea. He diligently furrows into the field he wishes to have an idea in, getting everything ready for a good idea when it comes. The idea appears all of a sudden, traditionally while he's outside the lab, walking or brushing his teeth, and immediately it grows, feeding on the mental bits and pieces he had been stockpiling for the purpose. As soon as the idea has fruited in his brain it begins to try and replicate itself into other brains.

The behavior of ideas as they jump from person to person has the same pattern as the spread of the Plague. Both are described by the same mathematical model. Under analysis, the booming exponential growth of scientific ideas actually breaks down into a series of recurring epidemics. Information, it seems, is a communicable disease; scientists, mere willing hosts. This is a handy model, because more is currently known about the broadcast of disease than about the workings of information nets.

There's a standard curve for the expansion of an epidemic—slow initial growth, rapid spread, and then tapering off as the population becomes saturated. In the past, scientists relied on intuition for signs of a groundswell in a new field, a hunch of something happening. These days it's important to find the most active areas soon, so the big daddy-os in government and private foundations (who generally aren't scientists) monitor the patterns of scientific papers carefully with computer-assisted programs and graph the resulting curve on the wall. When the curve begins to match the instep slope preceding a full-fledged epidemic, just at the point they figure a little shot of dollars will have maximum effect, they start diverting money in that direction and hope it takes off.

The computer-assisted program they use was developed by Eugene Garfield, who also started *Current Contents* in a converted chicken coop in New Jersey. The tool, called citation indexing, is the first mirror look at the texture of scientific information, a scientific method that examines scientific method.

Among the books in my university's library are several with photographs of Chinese scroll painting. Near the edges of the dark, moody paintings are the "chops" of previous owners of the scroll. These marks have become integral parts of the pictures, and, in fact, bring value to them. In the same way, the marks that an idea carries around, letting you know who else was infected with it, are called citations. They give extra value to the ideas. If a copy of the original was acceptable we could make several, add our own mark, and pass them along. We'd have a branching network of replicating ideas radiating out from the original, each one stamped with citations.

This is where the computer comes in. The computer keeps track of all the pieces being copied and passed around. That's the hard part, especially with several million citations per year. Each time a person mentions he's been infected with an idea, whether it's a recent one or an old one, it goes into the computer with two ends—one under the infectee, one under the infector.

The connections can now be traced both ways, back from an idea, or forward from an idea. Pick a paper at random from the middle of the network. The paper itself is marked with a bibliography that tells where the ideas were one step before this. As mentioned earlier, you can trace it all the way back to the beginning. Going the other way, the citation index traces where each branch of the idea travels afterward.

For instance, it would make a list of any book or paper in the future that acknowledges this text of mine as a source of ideas. A year from now you could follow all the divergent paths these ideas have led to, probably weird offbeat corners you would never have found otherwise. It's like being able to ask someone, "Who's playing with your ideas now?" If you can find the name of any idea, you can find where it came from, using references, and you can find where it's going, using citations.

With all these relationships burned into its brain, the *Science Citation Index*®, as it is properly called, can notice patterns that connect. The index comes either as a large 14-volume series, or in electronic text. Recently it has been taught to paint visual maps of the links within fields of research.

These are the maps of different neighborhoods in science. An experimenter's neighborhood is bounded by the klan of all those who wave to him and everybody he waves to, that is, the set of people who have acknowledged his ideas in their work as well as those he has cited in his work. Usually most of the people down the hall, or sometimes even working in the same lab, belong to a different neighborhood. The group tinkering in the laboratory next to ours is working in another galaxy, while a group in Sweden is our neighbor. The ones living closest together in a neighborhood are the ones bouncing the most ideas off each other. On the map this distance is measured out in units of citations. Nearness means a commonness of ideas.

You enter any neighborhood by dropping names, by citing an author. An individual scientist may be on the fringe of hundreds of communities, and near the heart of a few. In the center of every neighborhood is a nucleus of influential people who set the tone and pace, and occasionally bully the timid. In scientific communities these are the core people others concede as being personal donors of important ideas. A person new to the block doesn't need to unravel the tangled knot of relationships. Just a look at the map hints where the godfathers are. He could begin his research exploring the flock at the center and gather the fattest ideas there. Sweeping outward from the nucleus he would have to cast his net wider and wider, over more and more papers, to haul in the same catch of useful news. The hinterlands of one place, of course, could be the backyard of a different neighborhood.

The maps show only the major landmarks—folks everybody in town knows, and who know everybody else. The "empty" background is, in reality, a solid mass of other, smaller neighborhoods, hangout spots, quiet workers, a few hermits, and people moving around all the time. This is the map of our immediate neighborhood. We're not shown because we're mere peasant laborers in the field, but if we were, you'd see us living somewhere near the edge.

In fact, the map is changing all the time. The ground is not static, but an undulating throb. It's the same growing skin of ideas that foundation directors hold their breaths and run their hands over, searching for a sign of life on the charts. The tissue stirs, the pattern swirls. Maybe when the art is more developed we can make animated movies of ideas as they weave into colonies and every once in a while burst and contaminate everything with their strangeness.

So far we've got a fuzzy snapshot of the rough texture. It took zillions of citations reckoned by a hi-tech computer to verify what my neighbor scientists had been trying to tell me all along. The words they used, when they finally found them, seemed ridiculous, but were as close as any. The texture of scientific information, that is, the terrain of what we know, they kept saying, is a little thick in places, a little thin in others...like lumpy gravy.

BIBLIOGRAPHY

Durack D T. The weight of medical knowledge. *N. Engl. J. Med.* 298:773-75, 1978.

Garfield E. *Essays of an information scientist.* Philadelphia: ISI Press, 1977-1984. 6 vols.

Goffman W & Warren K S. *Scientific information systems and the principle of selectivity.* New York: Praeger, 1980. 189 p.

Kerkut G A. Choosing a title for a paper. *Comp. Biochem. Physiol. Pt. A* 74:1, 1983.

Thomas L. Hubris in science. *Science* 200:1459-62, 1978.

Tullock G. *The organization of inquiry.* Durham, NC: Duke University Press, 1966. 232 p.

Urquart D S. The use of scientific periodicals. *Proceedings of the International Conference on Scientific Information.* Washington, DC: National Academy of Sciences, 1958. p. 277-90.

Ziman J M. Information, communication, and knowledge. *Nature* 224:318-24, 1969.