

how much information

what kind of question?

engineering counts

i-school counts

what counts?

Quality of Information

Tuesday, Feb 8 2011

the ruck of history

And it is there, in the ruck of history, that they are able to locate an information explosion that means something substantive and which has discernible origins and contexts: that *these* types of information, for *those* purposes, for *those* sorts of groups, with *those* sorts of interest of are developing

--Frank Webster,

Theories of the Information Society, 2009

overview

recent history

context for Lyman & Varian, etc.

the longer view

Blair and beyond

looking for patterns

18th century

19th century

looking ahead

Year	# sci journal	# books, Johns Hopkins
mid 17c	2	
mid 18c	10	
1800	100	
1850	1000	
1900		100,000
1950	30-100,000	
1970		1,500,000

the recent ruck

1945 Bush, "As We May Think"

1960 Toffler, "The Quantity of Culture"

1962 Machlup - information/knowledge economy

1963 de Solla Price, *Little Science, Big Science*

1977 Porat - information sector

1973, 1979 Bell - information society

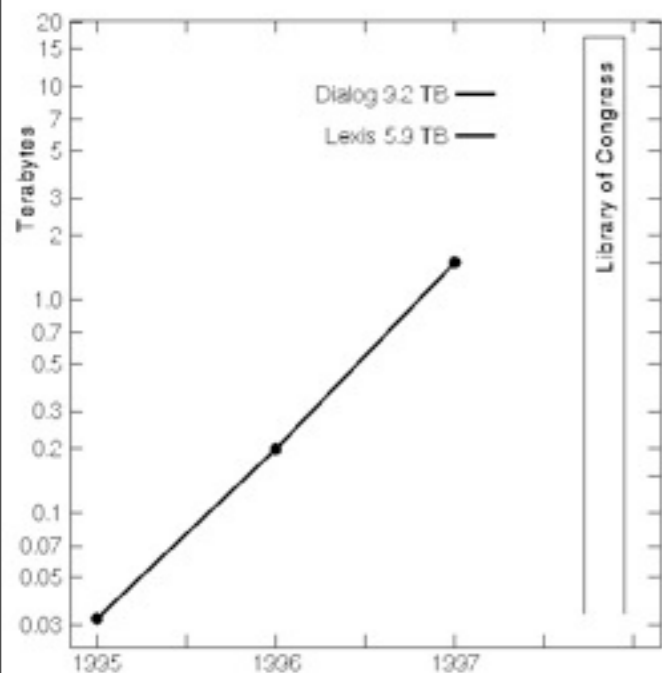
1996 Lesk - How Much Information in the World

1999, 2003 Lyman & Varian - *HMI*

2009 Bohn & Short - *HMI*

2010 Floridi, "the zettabyte era," "exaflood"

Web size



between Castells & Floridi

"Data from a large variety of sources can be combined in an equally large variety of ways with other data, thereby increasing substantially the amount of information that can be produced. ... the expansion is a systemic ... intrinsic characteristic of the contemporary world ... closely associated with sophisticated storage and updating mechanisms ... mutual implication of information with the technologies by which it is ... mediated ... [an] institutionally orchestrated game for obtaining information that is fresh and relevant results in information becoming readily depreciated and obsolete ... self-propelling, runaway .. expansion and growth"

--Jannis Kallinikos, *The Consequences of Information: Institutional Implications of Technological Change*, 2006

droning on

Military Is Awash in Data From Drones

By CHRISTOPHER DREW
Published: January 10, 2010

HAMPTON, Va. — As the military rushes to place more spy [drones](#) over Afghanistan, the remote-controlled planes are producing so much video intelligence that analysts are finding it more and more difficult to keep up.

still droning

In New Military, Data Overload Can Be Deadly



Doug Mills/The New York Times

At Langley Air Force Base in Virginia, teams monitor what they have nicknamed "Death TV" — live video from Afghanistan.

By THOM SHANKER and MATT RICHEL

Published: January 16, 2011

When military investigators looked into an attack by American helicopters last February that left 23 Afghan civilians dead, they found that the operator of a [Predator drone](#) had failed to pass along crucial information about the makeup of a gathering crowd of villagers.

 RECOMMEND

 TWITTER

back to the ruck

"All claims of this sort have their historical specificity, and one must always ask: who has to gain from assertions that information is chaotic, overwhelming, and out of control? The answer is usually found in the expert groups who offered solutions."

John Agar, *Government Machine*, 2003

who's counting?

toxic terabytes

"This year [2006], electronics manufacturers will produce more transistors ... than the world's farmers grow grains of rice ... four years from now, the world's information base will be doubling in size every 11 hours. ... kilobytes .. megabytes ... gigabytes ... terabyte .. petabyte .. exabyte ... zettabyte .. yottabyte"

"... when terabytes turn toxic ... no amount of disks will be enough to soak up the deluge"

"... taming the data beast ... data detox"

who's counting?

toxic terabytes

"This year [2006], electronics manufacturers will produce more transistors ... than the world's farmers grow grains of rice ... four years from now, the world's information base will be doubling in size every 11 hours. ... kilobytes .. megabytes ... gigabytes ... terabyte .. petabyte .. exabyte ... zettabyte .. yottabyte"

"... when terabytes turn toxic ... no amount of disks will be enough to soak up the deluge"

"... taming the data beast ... data detox"

footer 9

who's counting?

Financial support for HMI? research and the Global Information Industry Center is gratefully acknowledged. Our sponsors are:

AT&T

Cisco Systems

IBM

Intel Corporation

LSI

Oracle

Seagate Technology



HMI?
How Much Information?

problems & solutions

meet the 'net

'Virus' Hits Vast Computer Network

Philip J. Hilts Washington Post Staff Writer

The Washington Post (1974-Current file); Nov 4, 1988;

ProQuest Historical Newspapers The Washington Post (1877 - 1994)

pg. A1

'Virus' Hits Vast Computer Network

Thousands of Terminals Shut Down to Halt Malicious Program

By Philip J. Hilts

Washington Post Staff Writer

Thousands of computers at university and defense research centers throughout the United States were shut down yesterday and their links with other computers temporarily severed when they were hit by what appears to be the worst attack to date of a computer "virus," a program introduced by a saboteur.

By 3 a.m. computer systems linked to a research network called Internet began to fill up with extraneous material, slowing operations

and engorging memory files until the computers ground to a halt. Internet links as many as 50,000 computers, allowing users to send a variety of information to each other through the computers.

A computer virus is a program, or set of instructions, a computer programmer sends surreptitiously to other computers through telephone lines or by exchange of memory disks. The virus programs are introduced by piggybacking onto legitimate programs or messages. They may act immediately to carry out some malicious act, such as destroying data or overloading com-

puters, or may be instructed to hide until a set time.

A virus then copies itself and sends those copies to other computers where the chain of infection, action and further infection is continued. It appears that whoever introduced yesterday's virus did not intend to destroy files but merely to be a costly nuisance.

Researchers said it may not be possible to determine where the virus started. Asked who the saboteur could have been, Paul B. Pomes, senior research programmer in the computer services office

See VIRUS, A4, Col. 1

problems & solutions

A Continuing Series

Learning to Drink from a Fire Hose

Instant access to far-flung databases could soon be a reality, but how will we swallow a trillion bytes per day?



Last in a series

"Say I'm interested in a particular galaxy or star," says astrophysicist Larry Smarr, director of the National Center for Supercomputer Applications at the University of Illinois. "Ideally, I'd

like to tell the computer, 'Bring me all the data from any telescope and satellite that has ever looked at it. And then combine it into something useful.'" At a keystroke the information would just be there, he says, brought from anywhere in the world, filtered, transformed, integrated, imaged in vivid color, and ready for the user to explore—and all without the user knowing or caring where the data originally were.

Call this system the Digital Data Library. Neither Smarr nor anyone else actually has one—yet. But from all reports it's coming fast, conceivably as early as mid-decade. The technological pieces are already falling into place, especially with the spread of nationwide computer networks operating at higher and higher speeds. At least half a dozen laboratories, Smarr's supercomputer center among them, are now doing serious research into how to implement it. And perhaps most important, some kind of digital data library is fast becoming a scientific necessity.

"We're now in a time of large, interdisciplinary projects such as global change, biomedical research, even astronomy," says astrophysicist Barbara Mihalas, who is also interested in the data. "Ideally, I'd like to tell the computer, 'Bring me all the data from any telescope and satellite that has ever looked at it. And then combine it into something useful.'" At a keystroke the information would just be there, he says, brought from anywhere in the world, filtered, transformed, integrated, imaged in vivid color, and ready for the user to explore—and all without the user knowing or caring where the data originally were.

Doing research with this kind of data flow is going to be like drinking from the proverbial fire hose, says Mihalas, and massive computer assistance is clearly the only hope. At the same time, she adds, it's important to realize that better computer and network hardware is only one part of the solution. An even tougher challenge lies in getting scientists and administrators to change the way they think about data management.

At most institutions, for example, data management has traditionally taken a backseat to more urgent priorities such as buildings, people, and hardware. As a result, databases are not only dispersed geographi-

cally, but they are rarely indexed by anything resembling a national card catalog. "It comes like trying to find 'pineapple farmer' in an encyclopedia that doesn't have an index and isn't in alphabetical order," she says.

Under pressure of necessity, however, some of the big, centrally managed programs have begun to pioneer a more systematic approach to data management. Just last year, for example, as part of the legislation authorizing the Human Genome Project, Congress established a Biotechnology Information Center within the National Library of Medicine, which is part of the National Institutes of Health. The new information center, headed by David J. Lipman, was given the task of developing software and database systems for all of biology—including the human genome once it is sequenced. In parallel, the Human Genome Project office is now organizing a Joint Information Task Force to figure out how to handle massive databases that will be generated by the sequencing effort itself.

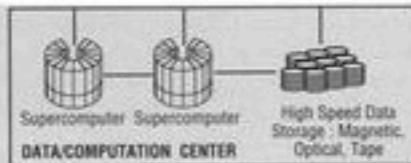
NASA, meanwhile, is setting up its Astrophysics Data System, which is intended to give astronomers access, through the NASA science network, to archived data from the Space Telescope and its other space astronomy missions—more than a dozen in all by the year 2000. Now scheduled to begin limited operation on 1 July, the system will also cover archives from all European and Japanese space astronomy missions.

As important as such organizational efforts are, however, they represent only one level of the conceptual change that's necessary to create a truly useful data library. An even deeper challenge arises from the fact that scientists—especially the most creative scientists—are very good at asking questions of the data that the database designers never thought of.

"Having this kind of flexibility is a very unusual requirement [to place on a database]," says James Ostell, software chief of the Library of Medicine's...

something to swallow

"I suppose ... a researcher in Idaho were to ask 'What do we know about the genetics of disease X?' A user knowbot...would interpret the request, determine how to go about answering it, and then create one or more messengerknowbots .. to appropriate the data .. more knowbots would integrate...arrange to display it in a meaningful way to the user himself--who never has to know about any of this"



overview

recent history

context for Lyman & Varian, etc.

the longer view

Blair and beyond

looking for patterns

18th century

19th century

looking ahead

the longer ruck



"It will soon be the employment of a lifetime merely to learn [books'] names. Many a man of passable information at the present day reads scarcely anything but reviews, and before long, a man of erudition will be little better than a mere walking catalogue."

Washington Irving, 1820

further back



"The most accomplished way of using books at present is twofold. Either, first, to serve them as men do Lords, learn their titles exactly and then brag of their acquaintance :-or, secondly, which is indeed the choicer, the profounder, and politer method, to get a thorough insight into the Index, by which the whole book is governed and turned, like fishes, by the tail. ... Thus men catch knowledge by throwing their wit on the posteriors of a book, as boys do sparrows by flinging salt upon the tail."

Swift, "Tale of a Tub," 1704



Seneca, 4 bc - 65 ad

further yet

De Tranquillitate Animi

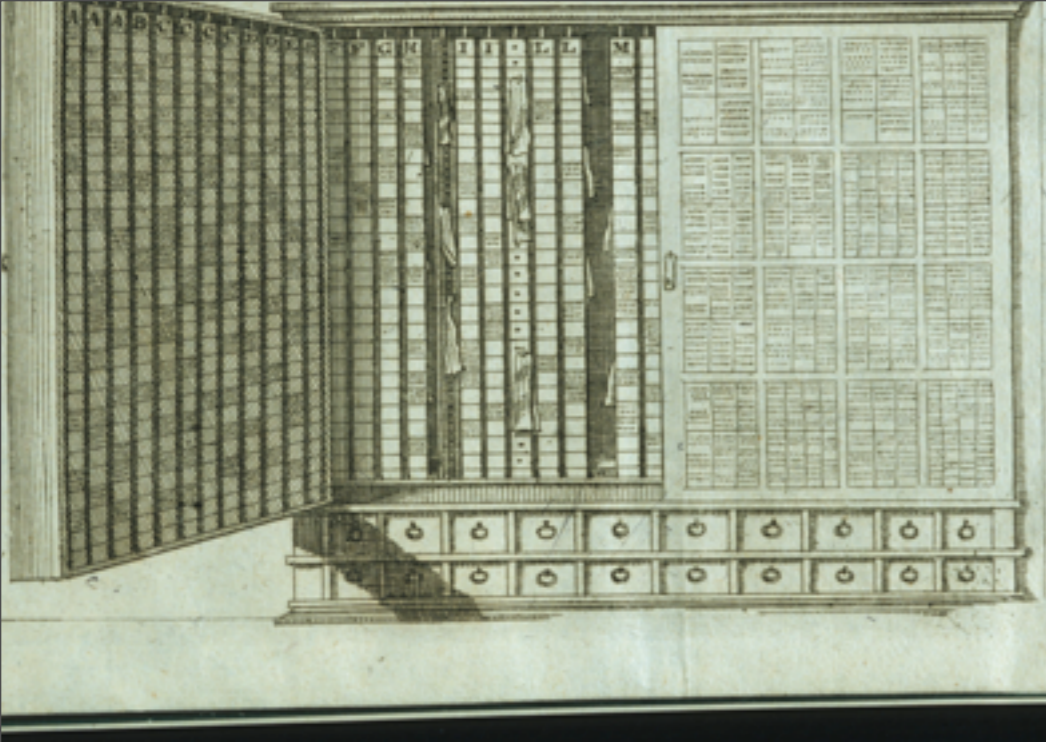
"Even for studies, where expenditure is most honorable, [developing libraries] is justifiable only so long as it is kept within bounds. What is the use of having countless books, and libraries whose mere titles their owners can scarcely read through in a whole life time? The mass of them does not instruct but rather burdens the student; and it is much better to surrender yourself to a few authors than to wander through many. Forty thousand books were burned at Alexandria; let someone else praise this library ... as did Titus Livius, who says that it was the most distinguished achievement of the good taste and solicitude of kings. There was no "good taste" or "solicitude" about it, but only learned luxury--no, not even learned, since they had collected the books, not for the sake of learning, but to make a show, just as many who lack even a child's knowledge of letters use books, not as the tools of learning, but as decoration for the dining room."

summing up?



"Ars longa, vita brevis,
occasio praeceps,
experimentum periculosum,
iudicium difficile."

Hippocrates, 460 bc



Vincentius Placcius
1642-1699
scrinium

continuity or change?

"selecting, sorting, and storing ...

"new genres ... dictionary, ... encyclopedic
compilation ... alphabetical index...

"reading in parts" ... consulted rather than read

"note taking ...

"Naudé's summaries ... abbreviations

"Isaac Newton .. unique in his practice
of dog-earing

"Gesner, cutting up his letters for reference."

Ann Blair, "Reading Strategies for Coping with
Information Overload ca 1550-1700," 2003

footer 18

overview

recent history

context for Lyman & Varian, etc.

the longer view

Blair and beyond

looking for patterns

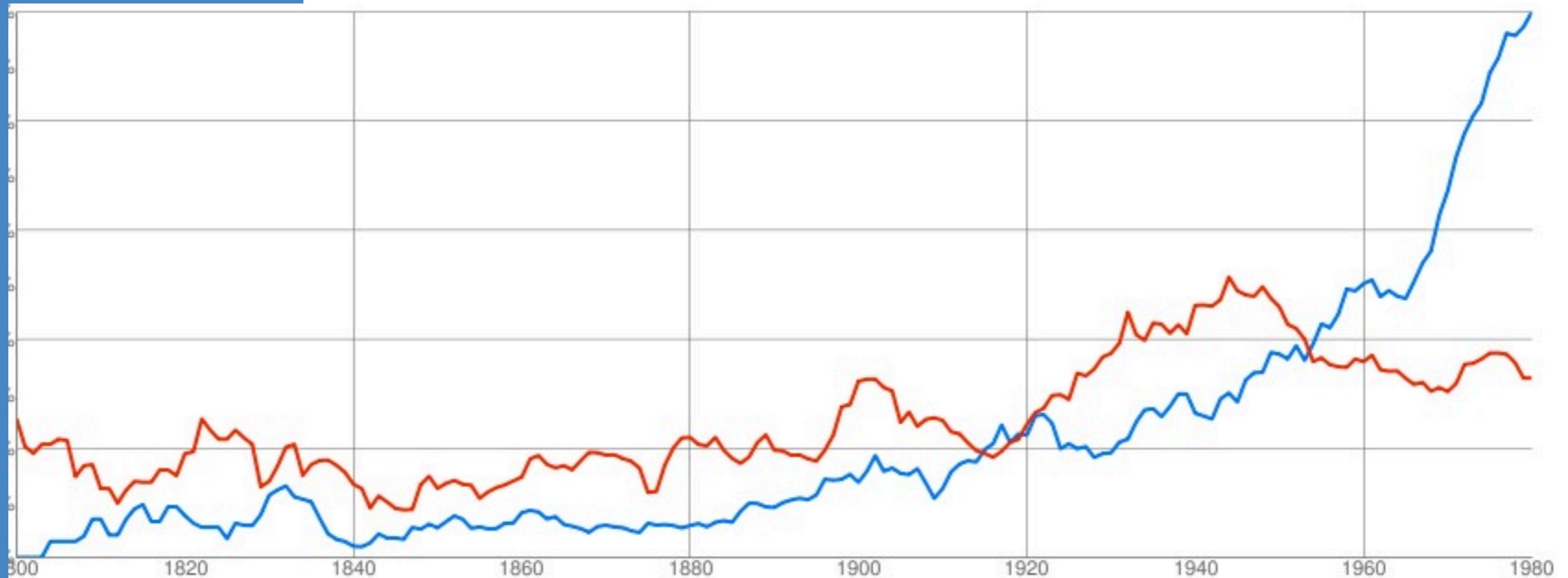
18th century

19th century

looking ahead

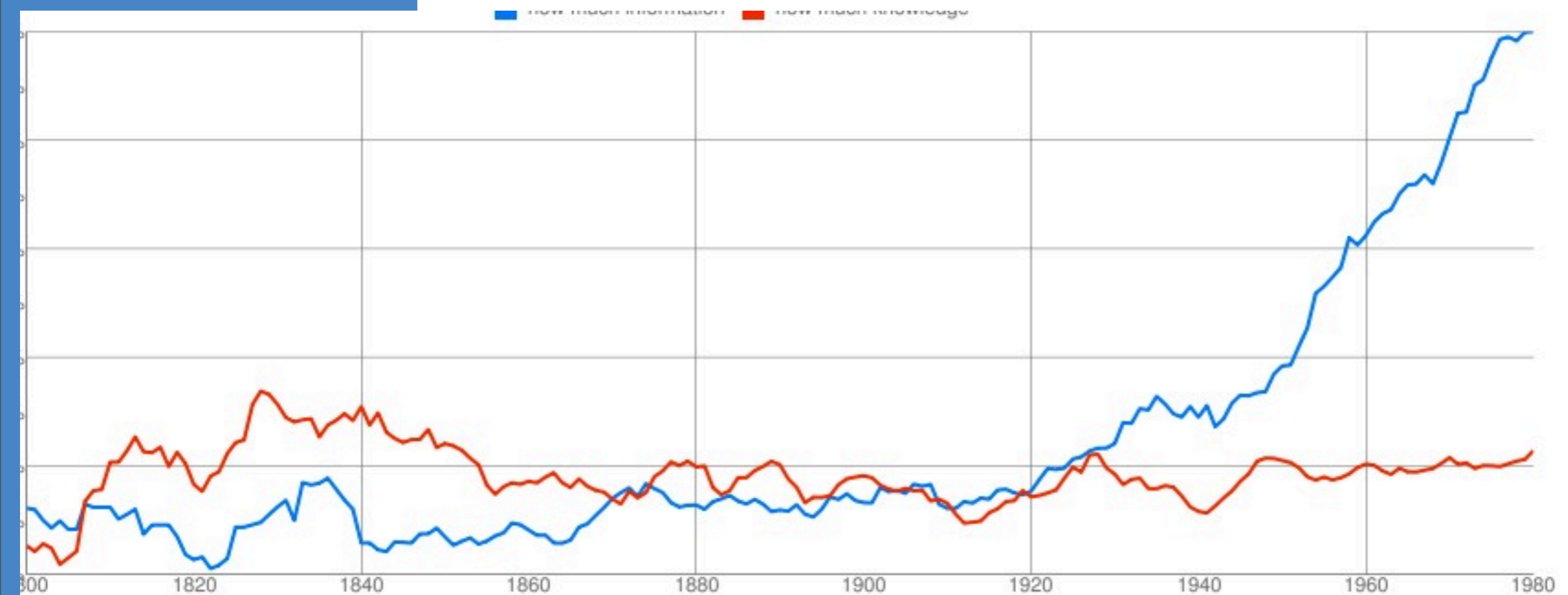
continuity or change?

too much information, too many books
1800-1980



continuity or change

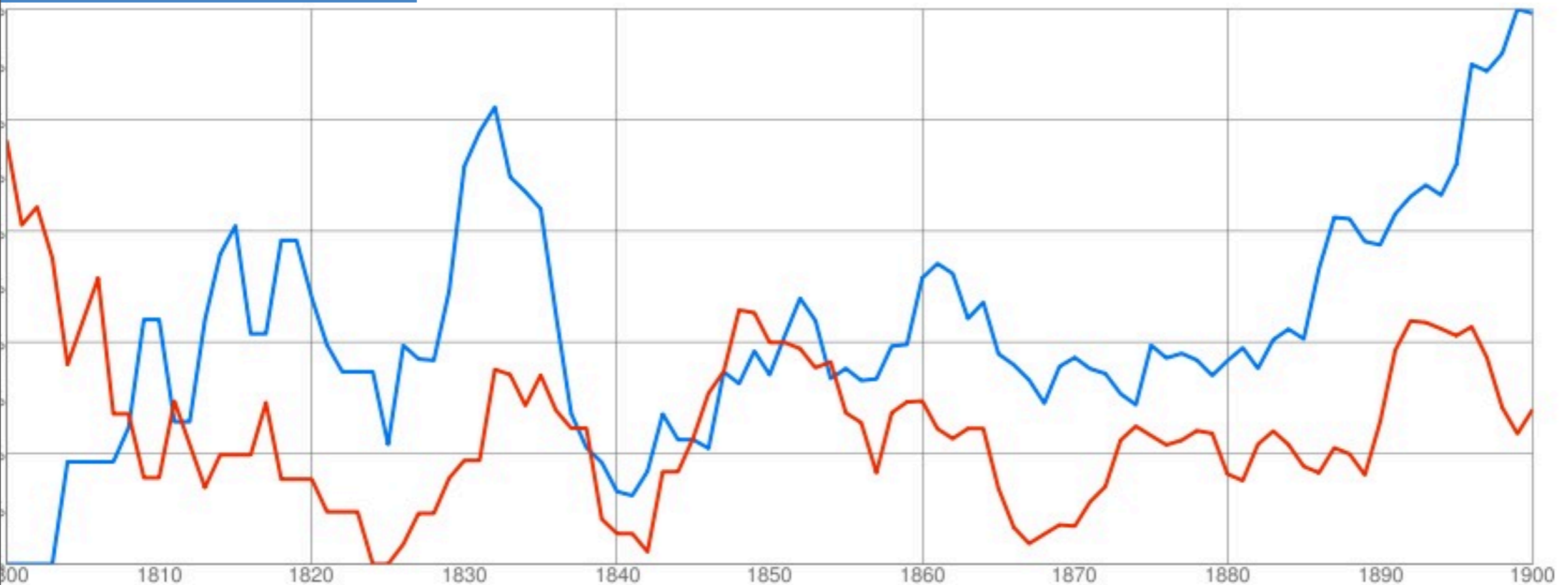
how much information, how much knowledge
1800-1980



continuity or change?

too much information, too little information

1800-1900



18th century

over to you

look at the data on the handout

[data from Gale's 18th century collections

do any patterns seem to emerge?

19th century

over to me

comparing

"The Register will be found to contain double
the quantity of information ever before
inserted in it."

Longworth, *Almanac*, 1799

"Johnson's *Journey to the Western Islands* ..
does not admit of abridgement, because every
sentence compresses such a quantity of
information."

Historical Magazine, 1799

footer 25

centuries of print



"The present age... may be styled, with great propriety, the Age of Authors; for, perhaps, there was never a time when men of all degrees of ability, of every kind of education, of every profession and employment were posting with ardour so general to the press."

Samuel Johnson, *The Adventurer*, 1753

trend lines?

The Reading Nation in
the Romantic Period

WILLIAM ST CLAIR

Number of titles printed in England:

William St. Clair, *The Reading Nation*, 2004

1630s	600	1700-50	500
1640s	1,600	1750-89	600
1650s	1,200	1790-1800	800
1660s	800	1800-1810	800
1670s	1,000	By 1827	1,000
1680s	1,500	("rising fast")	
1690s	1,400		

more selling

Times (London)

1784 -1900: **43** "more information"

7 ads (dictionaries, grammars, atlases)

25 books (reviews), newspapers, reports

7 education

1 telegraph

1 trade statistics

too much

[on newspaper tax]"The Times weighs generally more than four ounces ... this measure would [for taxing newspapers] affect The Times, while the other morning papers would be allowed to go free ... The Times, because it gave too much information to the country, ... would be obliged to pay more."

Times, 1855

still impossible

[the need for] "calm investigation ... For .. it is impossible to have too much information as to the social and economical relations of that agricultural population."

The Times, 1880

"The public, however, cannot have too much information at a crisis like this .. useful to give some account of the further Egyptian papers just issued."

The Times, 1882

footer 30

political debate

famine to glut

"If the government had erred at all, it has been in giving too much information."

The Times, 1885

really too much information

[re transatlantic cable]: "Every new ground of information, every earlier access to information .. is an addition to the trial and work of commercial life. Time was when a merchant in Liverpool or Glasgow received his business letters in a morning... then came the electric telegraph, which pursued him even into his evening circle ... evidence of overtaxed heads .. men have already more material of calculation than they can use A man may easily have **too much information** -- more than he can manage--one part of it modifying and qualifying another, till he does not know where he stands. But though every fresh addition, every earlier access to data is a trial of feeble, it constitutes the triumph of strong heads. They exult in the largeness, the multiplicity, the intricacy of their field of intelligence." --*Times*, 1866

footer 32

overview

recent history

context for Lyman & Varian, etc.

the longer view

Blair and beyond

looking for patterns

18th century

19th century

looking ahead

coming up

Thursday 10: no class

Tuesday 15: Rise of the Public Sphere

Habermas, Darnton

Tuesday 17th: Public Sphere exercise
suggestions welcomed