

Web 2.0, participatory media, and big data

LIB 1201

New York City College of Technology

September 23, 2013

What is Web 2.0?

What is (are?) Social Media?



What is Social Networking?

Baker on the Charms of Wikipedia

- Wikipedia began with public domain reference works
 - your thoughts/comments?
- Wikipedia’s “upper crust”—similar to traditional publishing, the most persistent content is contributed by a tiny percentage of editors
- Should Wikipedia have limits?

Jaron Lanier on Social Networking

While you arduously tend your fake self on Facebook, the company compiles a secret dossier about a more real you and everyone else so that access to you can be sold to political campaigns, teeth whiteners, or finance hucksters. You are the product, not the customer. Meanwhile the things you might offer online—your creative work, your opinions, your advice—are all made worthless in terms of the kind of real money that buys food and pays rent.

Jaron Lanier, *You are Not a Gadget Q&A*

<http://www.jaronlanier.com/gadgetcurrency.html>

Other Wikis

- [Wikileaks](#)

“a non-profit media organization dedicated to bringing important news and information to the public. We provide an innovative, secure and anonymous way for independent sources around the world to leak information to our journalists. We publish material of ethical, political and historical significance while keeping the identity of our sources anonymous, thus providing a universal way for the revealing of suppressed and censored injustices.”

- [Pbworks](#), [wikispaces](#), [wikia](#) all examples of wiki hosts

What is the big deal about Big Data?

- “Technology, decision-making, and public policy”
- The “land of the quants” – what?
- Surveillance potential
 - Privacy, commerce, security
 - Shopping for a deep-fat fryer online?
- “what does it mean to live in an era where things and people are infinitely observed?”
 - From Lohr, “[Sizing up Big Data](#),” New York Times Bits blog, 6/19/13

Tufte, Visual and Statistical Thinking

Success of Dr. Snow's detective work dependent on the following:

1. Contextualizing data
2. Making quantitative comparisons
3. Alternative explanations & contrary cases
4. Assessment of errors in data



WIKIPEDIA
The Free Encyclopedia

[Main page](#)
[Contents](#)
[Featured content](#)
[Current events](#)
[Random article](#)
[Donate to Wikipedia](#)
[Wikimedia Shop](#)

Interaction

[Help](#)
[About Wikipedia](#)
[Community portal](#)
[Recent changes](#)
[Contact page](#)

Toolbox

Print/export

Languages

[العربية](#)
[Bosanski](#)
[Česky](#)
[Dansk](#)
[Deutsch](#)
[Español](#)
[فارسی](#)
[Français](#)

Article [Talk](#)

Read [Edit source](#) [Edit beta](#) [View history](#)

Search

Big data

From Wikipedia, the free encyclopedia

Big data^{[1][2]} is the term for a collection of [data sets](#) so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage,^[3] search, sharing, transfer, analysis,^[4] and visualization. The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to "spot business trends, determine quality of research, prevent diseases, [link legal citations](#), combat crime, and determine real-time roadway traffic conditions."^{[5][6][7]}

As of 2012, limits on the size of data sets that are feasible to process in a reasonable amount of time were on the order of [exabytes](#) of data.^[8] Scientists regularly encounter limitations due to large data sets in many areas, including [meteorology](#), [genomics](#),^[9] [connectomics](#), complex physics simulations,^[10] and biological and environmental research.^[11] The limitations also affect [Internet search](#), [finance](#) and [business informatics](#). Data sets grow in size in part because they are increasingly being gathered by ubiquitous information-sensing mobile devices, aerial sensory technologies ([remote sensing](#)), software logs, cameras, microphones, [radio-frequency identification](#) readers, and [wireless sensor networks](#).^{[12][13]} The world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s;^[14] as of 2012, every day 2.5 [quintillion](#) (2.5×10^{18}) bytes of data were created.^[15] The challenge for large enterprises is determining who should own big data initiatives that straddle the entire organization.^[16]

Big data is difficult to work with using most relational database management systems and desktop statistics and visualization packages, requiring instead "massively parallel software running on tens, hundreds, or even thousands of servers".^[17] What is considered "big data" varies depending on the capabilities of the organization managing the set, and on the capabilities of the applications that are traditionally used to process and analyze the data set in its domain. "For some organizations, facing hundreds of gigabytes of data for the first time may trigger a need to reconsider data management options. For others, it may take tens or hundreds of terabytes before data size becomes a significant consideration."^[18]



A visualization created by IBM of Wikipedia edits. At multiple [terabytes](#) in size, the text and images of Wikipedia are a classic example of big data.

Tim Berners-Lee: [The Year Open Data Went Worldwide](#)

Reading/Blogging for Wednesday, September 25

Discussion: Copyright and fair use; open access

Reading: Center for Social Media, [The Code of Best Practices in Fair Use for Media Literacy Education](#), sections “Code” and “Principles” only

Viewing: Grey, Lessig, Faden

Assignment: Write one reading response blog post

Discussion Facilitators: Kimesha, Dimitri, Steve