

# Om å tenke stort

I løpet av få år er *big data* blitt et etablert uttrykk. Fenomenet vil trolig få stor medisinsk betydning. På norsk kalles det *stordata*.

Bruken av begrepet *big data* har eksplodert i takt med den massive veksten i tilgjengelig datamateriale og datakraft. Amerikanske informatikere brukte betegnelsen allerede i midten av 1990-årene for å beskrive datamengder som ikke kunne bearbeides av vanlige datamaskiner.

Selv om *big data* brukes hyppig som engelsk lånord i norske tekster, finnes det flere eksempler på oversettelser. Uttrykket *kjempedata* er brukt på regjeringens hjemmesider (1). Språkrådet anbefalte tidligere *massedata* (2). Likevel er det betegnelsen *stordata* som har sterkest vind i seilene. Stordata kom på listen «Årets ord» i 2014, og Språkrådet tror at ordet kommer til å fortrenge det engelske uttrykket (3). Jeg synes stordata er en god oversettelse og velger å bruke den.

## Stordata i medisinen

Uttrykket stordata beskriver både et fenomen og en fagdisiplin (4). Som fenomen omfatter stordata den enorme veksten i tilgjengelig data. 90 % av datamaterialet som var tilgjengelig på internett i 2013, ble generert i løpet av den foregående tiårsperioden (5). Den overveldende veksten i informasjon kan utvilsomt komme medisinen til nytte. Jeg vil nevne tre eksempler – tre ambisiøse prosjekter som har møtt hard kritikk fra skeptikere.

The Human Brain Project ble i 2014 del av EUs flaggskipsprogram, og prosjektet mottar dermed én milliard euro over en tiårsperiode. Målet er å simulere den menneskelige hjerne ved hjelp av superdatamaskiner som skal integrere enorme mengder av publiserte og upubliserte nevrovitenskapelige data (6). Ett av delmålene er å beregne virkningen av nye medikamenter direkte i datamaskinmodellen. Ved å studere kretsene i hjernemodellen håper man også å kunne predikere hva som går galt ved ulike hjernesykdommer. The Human Brain Project er altså en diger satsing, og kritiske røster har hevdet at målsettingene er diffuse og urealistiske (7).

Et annet medisinsk stordataprosjekt er IBMs «Watson». Dette er superdatamaskinen som slo alle amerikanske mestere i kunnskapsspillet Jeopardy i 2011 – en omarbeiding av IBMs «Deep Blue», som beseiret sjakkverdensmester Garri Kasparov allerede i 1997. Nå forsøker IBM å utdanne



The Blue Gene Q Supercomputer er en del av The Human Brain Project, her avbildet under en pressekonferanse på École Polytechnique Fédérale de Lausanne i 2013. Foto: Jean-Christophe Bott/NTB scanpix

«Watson» til superlege. Det endelige målet er at «dr. Watson» skal kunne lese normalt språk og gjennom vanlig kommunikasjon foreslå behandling og forutsi prognosen for enkeltpasienter. Resultatene fra dette arbeidet har begynt å komme, blant annet er IBMs storsatsing blitt brukt til å utforske over 70 000 enkeltstudier av det genregulerende proteinet p53 (8). Det ville vært umulig for én enkelt person å lese gjennom det enorme antallet studier. Basert på lesingen av disse studiene kom «dr. Watson» opp med flere forslag til nye proteiner som kan styre aktiviteten til p53. Nå brukes maskinen også i flere kreftprosjekter ved amerikanske klinikker (9).

Det tredje eksemplet er av litt mindre skala og handler om influensa. Google Flu Trends ble lansert i 2008 for å forutsi spredningen av influensaepidemier. Statistiske modeller brukes til å estimere influensaaktivitet ut fra søkeord i Google. I flere tilfeller har tjenesten imidlertid gitt feil anslag, men forskere fra University of California San Diego har nylig raffinert influensasøkeren (10). Fordelen med slike tjenester er at resultatene kommer i sanntid – konvensjo-

nelle epidemiologiske data blir som regel ikke tilgjengelig før etter at en epidemi har passert. Den hurtige tilgangen på informasjon gjør det mulig å planlegge målrettede tiltak mot epidemier tidligere.

## Stordata som fag

Stordata brukes også for å betegne en fagdisiplin der tradisjonelle metoder fra ulike vitenskapelige felter smelter sammen, spesielt er metoder fra statistikk, informatikk og matematikk viktige. På den måten minner stordata om andre motefagfelter som maskinlæring og kunstig intelligens (4). Mye tyder på at stordata kommer til å bringe medisinen store steg fremover.

### Mats Julius Stensrud

m.j.stensrud@medisin.uio.no  
Universitetet i Oslo

Mats Julius Stensrud (f. 1989) har mastergrad i statistikk fra Oxford University, studerer medisin ved Universitetet i Oslo og er tilknyttet Institutt for biostatistikk som forsker og underviser.

>>>

**Litteratur**

1. Syrstad RS. Kjempedata kan gi kjempevekst. Regjeringen 19.3.2014. [www.regjeringen.no/nb/aktuelt/Kjempedata-kan-gi-kjempevekst/id753334](http://www.regjeringen.no/nb/aktuelt/Kjempedata-kan-gi-kjempevekst/id753334) (6.2.2015).
2. Massedata. Termar frå Språkrådets dataterm-gruppe. [www.sprakrad.no/Tema/Terminologi-og-fagspraak/Datatermar](http://www.sprakrad.no/Tema/Terminologi-og-fagspraak/Datatermar) (6.2.2015).
3. Årets ord: framandkrigar/fremmedkriger. Språkrådet. [www.sprakradet.no/Toppmeny/Aktuelt/Arets-ord-framandkrigarfremmedkriger](http://www.sprakradet.no/Toppmeny/Aktuelt/Arets-ord-framandkrigarfremmedkriger) (6.2.2015).
4. Diebold FX. A personal perspective on the origin(s) and development of "Big Data": the phenomenon, the term, and the discipline. Social Sciences Research Network 26.11.2012. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2202843](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2202843) (10.2.2015).
5. Dragland Å. Big data – for better or worse. Sintef 22.5.2013. [www.sintef.no/home/corporate-news/Big-Data--for-better-or-worse/](http://www.sintef.no/home/corporate-news/Big-Data--for-better-or-worse/) (10.2.2015).
6. Markram H. The human brain project. *Sci Am* 2012; 306: 50–5.
7. Frégnac Y, Laurent G. Neuroscience: where is the brain in the Human Brain Project? *Nature* 2014; 513: 27–9.
8. Spangler S, Wilkins AD, Bachman BJ et al. Automated hypothesis generation based on mining scientific literature. I: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, NY 24.–27.8.2014. <http://dl.acm.org/citation.cfm?id=2623667> (10.2.2015).
9. Helping doctors to identify treatment options. IBM Watson. [www.ibm.com/smarterplanet/us/en/ibmwatson/implement-watson.html](http://www.ibm.com/smarterplanet/us/en/ibmwatson/implement-watson.html) (6.2.2015).
10. Davidson MW, Haim DA, Radin JM. Using networks to combine «big data» and traditional surveillance to improve influenza predictions. *Sci Rep* 2015; 5: 8154.

*Mottatt 6.2. 2015, første revisjon innsendt 13.2. 2015, godkjent 13.2. 2015. Redaktør: Marit Skaar Fjellhaug.*