

What is how much?
Misleading figures from Berkeley
István Dienes – Budapest

Senior researchers Peter Lyman and Hal Varian have recently published a study entitled „How much information?” [www.berkeley.edu] This study is an attempt to measure how much information is produced in the world yearly. The authors have studied several media and estimate annual „information production”, „information consumption”, „accumulated stocks of information”, rates of growth and „other valuables of interest”.

Any measurement which aspires to the rank of being qualified as „technical” or „scientific”, should have a well founded and well defined methodology. Information – the fundamental concept of classical theory of communications – assumes a transmitter and a receiver and a noisy telecommunication channel. There is no such amount there, as an immanent, absolute quantity of information „in a book” and as a thing as “production or consumption of information” *in classic theory*. [Braman] in her classical study collected several definitions to „information”, none of which is suitable directly for measuring „information” in an economic context. Furthermore, production, output, flow, consumption are internationally standardised macroeconomic concepts and quantities which are measured in the frames of theory and methodology of national accounting [SNA94]], and applied worldwide, even in the United States.

In the new Berkeley study, terminology is misleading and fundamental concepts are fuzzy. The authors refer to “stocks”, “flows”, but no true economic stocks and flows are studied. Lyman and Varian also try to measure „production of information”. However, information as understood in theory of communications is not a commodity, and no „information, as such” exists in statistics, as a commodity. Internet *gurus* and digital, *information age prophets* reflecting the interests of the infotainment industry keep preaching that “information” is an irregular commodity, e.g. “it does not vanish when consumed”. As a matter of fact, it is not an “abstract information”, or information in consumed status as when forming in one’s mind which is *the* commodity.. I can not consume a “bit”, or “Office 2000”. I can consume a CD, a copy of a book, etc. It is no “information” itself is a commodity, but certain objects, non-durable signals and services which carry/convey information, are commodities indeed.

The *authors adopt a “store model”*. As the authors write: “Soon it will be technologically possible for an average person to access virtually all recorded information. The natural question then becomes: How much information is there to store?” This approach is straightforward for a company like the storage media company which financed the study, or someone, who has a vision of a future information society with a monopolistic content distributor.

However, for a citizen/netizen or policy maker, this is not an issue. A store model has nothing in common with political economy of (information) commodities. A policy maker should be interested in economic efficiency and efficacy. The „production” and „consumption” figures of the authors can not be matched with costs figures. All policy issues are related to rarity, uneven distribution, *local* lack of commodities. It is an utopy and economic tohovahuhu to deal with quantities of originals *only, and*

instead of all copies. Billions of people just can not access the Net or can download only insignificant volumes from the surface, and never have production tools which make them capable to profit from those remarkable terabytes of originals. For instance, even in a European country, like Hungary, telecommunications companies may provide services at prices, which are higher than those in the richest countries, and not adequate to living standards. Millions never can step beyond “media information”, the kind of ambushes in Palestine, computer games, pop stars – the old content in a new(?) digital form.

In their „store modell”, Lyman and Varian *do not offer a generic definition* of their own, concerning their variables. The authors mistakenly adopted the information paradigm from the theory of communications and tried to merge it into macroeconomics. The result, is obviously a khymmera. Unless one created a new definition, should decide whether he - as an economist - is interested in the macroeconomy of production/consumption/accumulation of goods and services or, as an electric engineer is interested in transmission of signals in noisy channels. The elements of the two theories must not be mixed in some arbitrary ways.

One of the main conclusions of the study that while “information production” is growing exponentially, “information consumption” actually does not change. This is not surprising, given Lyman and Varian neglecting “machine information consumption”, i.e. consumption of information commodities by owners of machines, which commodities they input in (put into) their machines. Guess, what, if statisticians would try to meter electric-energy consumption, but neglecting boylers, refrigerators – and PC-s, measuring human electric consumption only?!. Machines, however, actually receive and consume more and more information, which is the far most important process of information economy and policy completely ignored by scholar world, including Varian. The more computers are installed, the more Winchesters, CD-s and floppies are not only produced, but, equally importantly, used and consumed. Ignoring consumption of floppies and Winchesters when accounting their production certainly distorts all kinds of statistics and conclusions.

Theoretical and methodological sloppiness, dull conceptual background is most outstanding in the chapter on broadcast media. Broadcast media are par excellence mass media. Surprisingly enough, the figures of production and consumption of mass communication are based upon broadcast time only, and do not take into account the billions of viewers/listeners. A mass medium without its audience, daily reach, would not be a mass medium at all. Actually, since communication of stations with their daily reach is accounted as if it were one channel with a single viewer, it is not surprising, that phone conversations were found to be more productive than mass media. Furthermore, the authors are not sure whether radio broadcasting and print media should be measured in compressed or uncompressed bytes. For a statistician, it is clear that statistics should reflect facts. It is a matter of fact, that broadcasting stations transmit programmes, and do not transmit compressed programmes.

No attempt was made to provide figures for Internet downloadings – “production of contents” by copying originals.

Statistics is a science about measuring, observing reality, „as is”. A scholar with social responsibility do not publish sensational – but apologetic or misleading –

figures. Internet enthusiasm coupled with and supported by surveillance societies, which protect copyright of (softwares and databases of) big producers, but does not really protect (copyright and privacy of e-mails, browsing paths and Winchester drives of) people does not help nations to meet challenges on the new era. Assuming a „monopolist world distributor”, a centralised distributing system for „information” is not real, our living societies obviously are not centralised.

Furthermore, the study does not deal with us, *human individuals*, whose information storage and processing capability with some 10^{12-14} neural cells, and 10^3 links each, in each of us, is still much more than anything mentioned in the study. *Humans* should stand in the centrum of any serious worldwide quantitative economic model and theory of information society.

Methodological inadequacy and sloppiness have lead the authors to the assumption of „democratisation of „data””. As far as no good data are available for such key-importance indicators, as production of data by individuals on PC-s, there is no solid basis to declare such a thing.

„Production, consumption, use, accumulation of information” has been defined as production, consumption, use, accumulation of information commodities in [SNIA], a much earlier, but better-founded Berkeley-Budapest attempt to define a satellite account to SNA, in value *and* natural-unit terms (bits).

This study, when lacking conceptual clarity and scientific rigor, drawing invalid conclusions and avoiding real problems, moved the Economist’ layman-columnist to write: „Exactly 3,344,783 megabytes. So you know why you’re always out of breath.” My question is this: what is exactly 3,344,783 megabytes? Nations need and demand real figures which reflect information activity of their own, governments, enterprises, various big brothers in economic aspects, real economic flows and knowledge stocks, standard figures by official statistical institutions. Its time statistical offices woke up and prevent laymen and digital prophets to blow their trumpets. This will be real democratisation of data.

[Braman] Braman S. (1990) Defining Information: An Approach for Policymakers
Telecommunications Policy 13(3) pp 233-42

[SNA] System of National Accounts (1993)
Inter-Secretariat Working Group on National Accounts, Eurostat, IMF, OECD,
UNSO, World Bank.

[SNIA] Dienes I. (1994) National Accounting of Information. Reference Manual of
SNIA, Version 1.1
Manuscript, Berkeley, Budapest <http://free.x3.hu/infostat>