Surfing versus drilling in fundamental research

In 1633, as Galileo left the courtroom where his scientific opinions, formed by drilling deep into astronomical data, had been investigated by the Inquisition, he mumbled softly: “Eppur si muove” or “And yet she is moving”, i.e. the earth. He had been coerced to confirm the “alternative” fact or theory that the earth does not move and that the sun orbits the earth, this to avoid death at the stake. The current surge in the emergence of alternative facts or theories, spread through surfing on the internet, yet being at odds with the truth, is nothing new. It only has less deadly consequences these days.

In research, questions or hypotheses are being formulated that are to be answered or confirmed/rejected based on facts and logic. Confronted with data obtained through research or gathered in the internet, one has to determine whether the answers and proofs found are reliable. Here four aspects are of relevance:

1. Are there any unproven assumptions underlying the research, data or theory? How will the approximations used in the research, or inherent to a model or theory, influence the results obtained?
2. Is there sufficient statistics, i.e.a sufficient number of independent observations, to draw conclusions? Are the results unreliable due to undersampling, i.e. non-representative sampling? How large is the uncertainty (error) due to assumptions, approximations and poor statistics or sampling?
3. Are there any hidden or confounding variables, i.e. factors that were not considered in the research, model or theory, but may influence/determine an observed correlation?
4. Causality? Is there an explaining mechanism for an observed process or correlation?

When searching for information on the internet, the data found are often insufficiently documented to answer these four questions, with the consequence that one cannot determine their reliability. In other words, the internet has a limited value when one wishes to drill deep.

In my research, the major use of the internet is to find literature regarding a specific topic, and to look up basic physical and mathematical data and formulae. Even translational dictionaries available on the internet, which I sometimes consult to find Dutch/English or Dutch/German translations of sayings or expressions, still perform poorly compared to a high-quality dictionary: Many complex language phrases and expressions are not properly translated.

Fundamental research is impossible without drilling deep. Yet surfing has its value when hitting upon analogous data patterns, models or theories in other fields of science. These may induce ideas for solutions to problems in one’s own field of interest. This means that surfing should be done, but only rather limited in time, at least compared to drilling.

It is a popular thought that big data in itself can generate new scientific insight. Yet this is doubtful because the availability of lots of data does not guarantee any correlation or underlying mechanism to be abstracted from it. Figure 1 illustrates this by describing the various steps taken when formulating a model at the atomic or molecular level to simulate bio-molecular behaviour on a computer. Although the bio-molecular data available in the internet is huge, one is often limited to only producing pretty pictures without real sci-
Scientific content due to the considerations of step II in the figure. This state of affairs is also responsible for the limited predictive power of computational models in biology, at least compared to their application in some other fields of science, see Figure 2.

Although pretty pictures without scientific content may be appealing, the tendency to value an image of research or of a research organisation higher than their contents or the truth constitutes a threat to drilling research. This tendency is detrimental for a healthy development of academic institutions, but no cause of real worry as long as the number of deep-drilling researchers stays much larger than the number of surfing ones.

Figure 2

Computational Science and Engineering: use of computer modelling or simulation
Areas with a computational branch

What about biology?
- complexity of systems (space, time)
- inaccurate description of fundamental interactions
- fragmented data on a great variety of systems
- black-box character: missing view on causality

Prof. Dr. Wilfred van Gunsteren
ETH Zurich
Laboratory of Physical Chemistry
HCl G237
8093 Zurich
Switzerland
Phone +41 44 632 5501
wfvgn@igc.phys.chem.ethz.ch
ORCID 0000-0002-9583-7019
http://www.igc.ethz.ch

Citation: van Gunsteren WF: Surfing versus drilling in fundamental research. Infozine 2018, Special Issue 2,18–19. DOI 10.3929/ethz-b-000294373
Copyright: Wilfred F. van Gunsteren, CC BY NC ND 4.0
Published: November 15, 2018