Big data and the emergence of new ‘dissipative’ structures

Daniel Pauly*

Institute for the Ocean and Fisheries, University of British Columbia, Vancouver, BC, Canada

ABSTRACT: This essay suggests that humanity has experienced several instances where lots of information (‘big data’) had to be accommodated, which led to new structures for channeling the subsequent data flows. These structures, such as articulated speech and writing, would be analogs to the ‘dissipative structures’ that emerge in physical systems characterized by strong energy (i.e. heat) gradients. Additional examples from oceanography, meteorology and ecology are given, with some emphasis on the prescient work of Alexander von Humboldt, whose identification of ecological communities was based on the occurrence records of multiple species. His lead was initially not followed up, but it can be now, as millions of occurrence records are available, along with the technology to manipulate them. The structures that will emerge in the process, however, are as unpredictable as dissipative structures in physical systems.

KEY WORDS: Information transfer · Language · Data sharing · Humboldt · Ecology · Occurrence records · Aquamaps

The term ‘big data’ was essentially unknown prior to the 1960s (Fig. 1), although several scientific disciplines were then already blossoming that produced and used huge quantities of data, and had addressed, or even resolved, the associated issue of data sharing. It appears that the driver for the emergence of human speech was the need to keep track of social interactions in increasingly large groups of people (Dunbar 1998), and that the driver for the emergence of writing was the need to keep track of increasing numbers of commercial transactions (Lieberman 1980). Thus, it can be argued that language and writing were new structures created both for and by massive information transfers.

Similarly, the expansion of the European horizons in the Age of Discovery required a method to name the many animals and plants that were brought back from far away, beyond the 500 or so taxa that ‘folk taxonomies’ can usually handle (Berlin 1992), and thus the importance of the Linnaeus (1758) binomial and hierarchical system, which accommodated (and still does) an ever increasing terrestrial and oceanic biodiversity, in spite of various challenges (Boero 2010).

Thus, the pattern may be that more data or information lead to new structures to assimilate these data or express the information they contain.

This pattern may hold with texts incorporating lots of ideas, e.g. encyclopedias. One of these, the massive Encyclopédie of Diderot and D’Alembert

© The author 2017. Open Access under Creative Commons by Attribution Licence. Use, distribution and reproduction are unrestricted. Authors and original publication must be credited.

Publisher: Inter-Research · www.int-res.com
A fourth element is usually required (and available) for biological specimens: the person who has done the sampling, which allows connecting the specimens in question to the scientific literature (Froese & Pauly 2013).
graphy are now converging. Their current and historic data sets (the latter enriched by rigorous data recovery programs, e.g. for atmospheric and oceanographic data gathered by the Axis powers during World War II) are jointly run for both short-term predictions of the weather and long-term predictions of the climate (Edwards 2010). Here, ‘big data’ not only created new patterns, but led to the emergence of programs of actions to undertake, or to ignore, at our own peril.

Big data may also help to overcome some of the divisions between the humanities and the sciences, e.g. through the introduction of quantitative approaches to study phenomena that have so far been approached phenomenologically. Examples are the study of ‘Ngram’ in millions of scanned books (see Fig. 1 and Michel et al. 2011, Stergiou 2017), or the construction of thousands of trees, and the selection of the most likely to depict the evolution of languages (Gray & Atkinson 2003), and even of creation and other myths (d’Huy 2016).

Until recently, ecology had no standard protocols for sharing data and no culture encouraging the practice, hence the frequent exhortations in leading scientific journals for more data recovering and sharing (e.g. Griffin 2017). However, ecology will eventually catch up with big data. Notably, it is likely that the hundreds of millions of occurrence records in the taxonomic literature and in museum collections will be retrieved by artificial intelligence programs. This would allow for better following up on biogeographical ideas such as those of Alfred Wallace (Barber et al. 2000). It would also allow for improving the extent and quality of the coverage of existing initiatives, such as OBIS, and thus for improving derived products, such as Aquamaps (Kaschner et al. 2008; www.aquamaps.org), which link these records with environmental parameters (temperature, depth) to generate probabilistic maps of the distribution of various marine and freshwater taxa (see e.g. Fig. 2).

The recovery of a massive number of occurrence records would also enable us to follow up, albeit belatedly, on Humboldt’s ideas, and to track the effects of global warming on the distribution of communities of organisms. This topic is still in its infancy because many biologists persist in dealing with global warming one species at a time, despite concepts and approaches being available for dealing with ensembles of species (see e.g. Cheung et al. 2010, 2013).

In physics, there is an analog to the above contention that massive data or information lead to new

\[ \text{Indeed, we are the first civilization that will be able to predict its own demise (see Oreskes & Conway 2014)} \]
structures to assimilate these data or express the information they contain. This analog relates to the ‘dissipative structures’ (Nicolis & Prigogine 1977) that emerge when energy gradients become so strong that energy is not transferred by a linear increase of the mechanism used when the gradients are weak. Such dissipative structures emerge spontaneously in pots of boiling water, or as the Hadley (wind) cells that transfer heat from the tropics to the poles. Indeed, life itself may be a dissipative structure, as well.

Here, I simply contend that massive data create the structures through which they are processed and will flow, and that the shapes and dynamics of these new structures are not predictable from the shapes and dynamics of the structures that accommodated the smaller data flows.

There is a long tradition of old men predicting a future that they will not experience and that mostly does not turn out the way they predicted. However, this author — also an old man — predicts that we cannot predict what big data will create in the longer term, in any scientific discipline and in society at large, good or bad.

LITERATURE CITED


Linnaeus C (1758) Systema naturae per regna tria naturae, secundum classes, ordinus, genera, species, cum characteribus, differentiis, synonymis, locis. Tomus I. Editio decima, reformata. Impensis Direct. Laurentii Salvii, Holmiae


Editorial responsibility: Konstantinos Stergiou, Thessaloniki, Greece

Submitted: April 25, 2017; Accepted: June 25, 2017

Proofs received from author(s): August 15, 2017