

Book Review: Life Out of Sequence: A Data-Driven History of Bioinformatics. Chicago: University of Chicago Press. 63£ Hardback/20.86£ Paperback. 304 Pages.

We are living in a moment of overwhelming hype over everything about 'big data', when promises of overarching understanding and unlocking the secrets of life are again littering mass media and scientific publications alike. At such a time, the emergence of ethnographic and historical studies of data practices is invaluable. Understanding the roots of data-intensive science and its current manifestations is key to unveiling the opportunities disclosed by these methods - and the limits of their reach. Stevens' book is one of the first of such studies to be published, and it proves tremendously useful in a variety of ways: it contains an excellent account of the spaces, labour and priorities characterising the development of the Broad Sequencing Centre at MIT and the Ensembl database at the European Bioinformatics Institute; traces a history of the relations between biology and computer science, and specifically of the development of sequence databases, genome browsers and heat maps; pays attention to both the scientific and the socio-political issues involved in the choice of coding, software, formats, annotation systems, databases, ontologies and visualisation tools used to handle data; and emphasises the relation between technological innovation and the economies of data production that characterise the world of genetic sequencing. It also makes for an engaging read, as Stevens' skillful writing shines light on data practices with clarity and wit - a real feat, given the highly technical nature of some of this material.

For all its merits, this book's main weakness is also what makes it such an interesting document of its time: a strong faith in and focus on the power of new technologies - and particularly computing - to change biological research in radical ways. In fact, Stevens argues that biology has brought nothing to computing: the marriage of the two disciplines has generated a one-way flow of methods and ideas from computing into biology, and not vice versa. The argument is well made, and Stevens provides empirical evidence for his claim. However, the evidence is limited to a few highly visible cases in cutting-edge, well-funded American and British molecular biology, which is a very specific interpretation of the 'data-driven' history promised by the title. Further, it is not defended in relation to a broader understanding of the life sciences and/or computing, both in terms of geographical spread and in terms of areas of specialisation involved.

The title of the book accurately reflects this perspective: bioinformatics is interpreted through the lens of the use of computing technology in (the most visible initiatives in) molecular biology, and whether this is representative of biology or even bioinformatics as a whole remains an open question, which is not confronted here. This is a reasonable way to structure a book that is already bulging with information and sharp insights. And yet, the question of representativeness becomes problematic when considering some of Stevens' claims, which are cast clearly and provocatively with a general scope. For instance, the idea that computers imported ready-made reasoning from physics

in biology – an argument that may be defensible in the case of genomics, but becomes unsustainable in the face of the serious challenge posed by the digitalisation and analysis of data generated in other areas of biology, and even of other ‘omics’ such as metabolomics. A related claim is that data are themselves defined by computation. This is taken as a given in the book, and raises the question of why many biologists continue to generate and use data that are difficult to compute (such as photographs, field observations, samples and even specimens themselves), and indeed regard the analysis of such data as crucial to their work.

I find it important to note that different branches of biology adopt computation not only at different rates, but in entirely different ways. I am sympathetic to Stevens’ idea that computational approaches to biology may become so ubiquitous that ‘bioinformatics’ will disappear as a meaningful term of reference, and yet I think that more attention should be paid to the organic intertwining of biological and computation techniques, the varieties of informatics that are being developed in association to the specific needs of different areas of the life sciences, and the importance of familiarity with biological objects and processes when interpreting digital data (a topic which is already investigated in the work of James Griesemer, Werner Callebaut, Edna Suarez and myself, among others). The results of such research may run counter to Stevens’ final argument about the revolutionary nature of what he calls Biology 3.0, in which “biological objects and their informatic representations will become apparently interchangeable” (p.219). Overall, because of its focus on technology, this book tends to pay little attention to data as epistemic and scientific objects, and particularly to their biological provenance and its relevance to interpretation. Nevertheless, Stevens’ volume certainly constitutes a required read for anybody interested in data-intensive biology, and particularly its analysis of the emergence of bioinformatics as a discipline represents a key contribution to historical and sociological understandings of this area.

Reviewer: Sabina Leonelli, Exeter Centre for the Study of the Life Sciences (Egenis) & Department of Sociology, Philosophy and Anthropology, University of Exeter

s.leonelli@exeter.ac.uk