Small data, big ideas

Urgent decision-making can't wait for big data

Iadine Chadès[1*], Sam Nicol[1]

[1] CSIRO, Dutton Park, 4102 Queensland, Australia

Corresponding author: iadine.chades@csiro.au

Big data problems have attracted an increasingly large proportion of private and public sector funding, influencing research efforts worldwide[1-4]. We welcome the exciting opportunities to gain new insights from big data. However, the research community, funding agencies and the private sector must think critically about how this rapid shift will affect our ability to solve pressing problems for which we do not have ways of obtaining big data, yet urgently require decisions to be made in a fast-changing world. In 2015, the US National Science Foundation awarded funding to five times more big data projects than in 2012; an increase in funding of $100 million in just three years[1]. In 2014, Wlodarczyk and Hacker [5] reported a ten-fold increase in big data publications over the last three years. Here, we discuss two situations where big data fails to address major problems facing society: where resources are limited but decisions must be made, and where big data collection is not possible. In both cases, we argue that solutions should come from intelligent use of small data, driven by decision science approaches.

The resources spent collecting big data outweigh its benefits. Collecting data consumes two of our most limited resources: time and money. Investing in collecting more data to identify best actions to protect threatened species has repeatedly been shown to be an inefficient use of funding[6] that could have been invested in active, on-ground actions. In the most dramatic case, an over-emphasis on monitoring can delay the decision-making process so severely that critical objectives cannot be met (e.g. species lost to extinction; invasive species establish and spread)[7,8]. Another limitation of collecting big data occurs when confronted with rapidly changing complex systems such as those that depend on climate variability[9,10]. In these cases more data does not always lead to higher predictability; for example where novel ecosystems emerge rapidly and unexpectedly from small increases in pressures[11]. Decision science methods such as value of information analysis can be used to rapidly evaluate the expected gains of learning, which can inform whether big data collection is a worthwhile investment[12-14].

Sometimes collecting big data is simply not feasible. Decision-makers are often required to make decisions with no or little information. This is the case with unexpected or emerging issues for which data does not exist in the relevant context. For example, in natural or agricultural systems, a fast response to a new invasive pest or disease outbreak can make the difference between a successful crop and eradication program or decades of costly management and quarantine programs[15,16]. Similarly, in ecology, there are situations where population sizes and detectability are too low to create big data sets (e.g. low abundance and difficulty to detect our most threatened and cryptic species[17]). Often the only data available comes from a relatively small pool of domain experts[18] that constitute our only chance of saving our species[19,20], or predicting the next likely locations of an invasion[21]. Other problems in domains such as health, defence, and social-sciences may also be constrained by the lack of big data options either because these big data do not exist or collecting them violates human ethics[22,23]. In many of these disciplines, decision-scientific approaches such as structured decision making and adaptive management can provide intelligent ways to use small data sets to make the best decisions given existing data[24-28].

For some of the world's most pressing problems, big data does not provide the answers. Smart use of modelling and development of small-data methods remains the best approach. An over-emphasis on big data funding to solve all our problems forces many of our brightest scientists to step away

from our most urgent problems for which, more than ever, big ideas are needed to solve our small data problems.

Acknowledgement

We thank C. Fletcher, D. Metcalfe, A. Tulloch, G. Iacona and T. Chams for commenting on previous versions of this manuscript, and Fred Reverchon for the title.

References

1 Huang, Y., Zhang, Y., Youtie, J., Porter, A. L. & Wang, X. *PLoS ONE* **11**, e0154509,(2016).
2 CSIRO. (CSIRO, 2015).
3 Mervis, J. *Science* **336**, 22-22,(2012).
4 LaRiviere, J., McAfee, P., Rao, J., Narayanan, V. K. & Sun, W. *Harvard Business Review* (2016).
5 Wlodarczyk, T. W. & Hacker, T. J. *International Journal of Big Data Intelligence* **1**, 172-180,(2014).
6 McDonald-Madden, E. *et al. Trends in Ecology & Evolution* **25**, 547-550,(2010).
7 Martin, T. G. *et al. Conservation Letters* **5**, 274-280,(2012).
8 Martin, T. G., Camaclang, A. E., Possingham, H. P., Maguire, L. A. & Chadès, I. *Conservation Letters*, n/a-n/a,(2016).
9 Watson, J. *Nature* **534**, 437-437,(2016).
10 Nicol, S., Fuller, R. A., Iwamura, T. & Chadès, I. *Proceedings of the Royal Society of London B: Biological Sciences* **282**,(2015).
11 Hobbs, R. J., Higgs, E. & Harris, J. A. *Trends in Ecology & Evolution* **24**, 599-605,(2009).
12 Schlaifer, R. & Raiffa, H. *Applied statistical decision theory*. (Clinton Press, Inc., 1961).
13 Canessa, S. *et al. Methods in Ecology and Evolution* **6**, 1219-1228,(2015).
14 Runge, M. C., Converse, S. J. & Lyons, J. E. *Biological Conservation* **144**, 1214-1223,(2011).
15 Simberloff, D. *Conservation Biology* **17**, 83-92,(2003).
16 Ward, M. *Food Security* **8**, 77-81,(2016).
17 Chadès, I. *et al. Proceedings of the National Academy of Sciences* **105**, 13936-13940,(2008).
18 Martin, T. G. *et al. Conservation Biology* **26**, 29-38,(2012).
19 Chadés, I. *et al. Conservation Biology* **29**, 525-536,(2015).
20 Firn, J. *et al. Global Change Biology* **21**, 3917-3930,(2015).
21 Martin, T. G. *et al. Biological Invasions* **17**, 3197-3210,(2015).
22 Vayena, E., Salathé, M., Madoff, L. C. & Brownstein, J. S. *PLoS Comput Biol* **11**, e1003904,(2015).
23 Wilson, S. *Nature* **519**, 414-414,(2015).
24 Gregory, R. *et al. Structured decision making: a practical guide to environmental management choices*. (John Wiley & Sons, 2012).
25 Walters, C. *Adaptive management of renewable resources*. (1986).
26 Weber, M. *European Journal of Operational Research* **28**, 44-57,(1987).
27 Chades, I. *et al.* in *Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI-12).*
28 Chadès, I. *et al. Theoretical Ecology*, doi:10.1007/s12080-12016-10313-12080,(2016).