

# Good data, bad data and ugly data

Researchers can be expected to employ a vast range of experimental techniques in pursuit of a scientific question. Making efforts to seek expert advice and develop the competency to generate, store and analyse high-quality data when first using an approach will save time in the long run.

Where previously it was possible to build a career mainly focusing on a relative few of the many experimental disciplines that might conceivably be employed in our field, today's microbiologists may find themselves having to generate and handle a far wider range of data types — from molecular biology, biochemistry and cell biology, through animal husbandry and surgery, to computer programming, bioinformatics, structure determination and statistics, with a multitude of others beside. While this certainly leads to a broad general experimental understanding, it risks a generation of researchers that are Jack or Jill of all trades, yet master of none. It is also not unheard of that the principal investigator(s) leading a project do not themselves have deep experience in a particular data type important for the project, which together can limit the quality of oversight that they can provide for any single project researcher. Furthermore, research projects are commonly carried out in time-limited situations, with deadlines imposed by the need to publish before a position ends, to find a next post, to apply for fellowships and funding. A lack of experience to draw upon when designing and undertaking a series of experiments, as well as the underlying time pressure, can lead researchers to cut corners, whether knowingly or not. If we are not careful, practices can creep in that lower the quality, usability and reliability of data and the analysis built thereupon.

Even for standard techniques, problems can begin at the experimental design phase when considering what technical or biological repetitions are feasible, which positive and negative controls can or should be included, how the samples should be processed, and the data generated, analysed and stored. Without in-depth experience in a particular approach, it is easy to make a misstep that undermines an entire experiment. In addition, the technology used to generate, process and analyse both visual and non-visual data types changes with time, providing increased ease of use, finer resolution or greater volumes of data to be incorporated, but also poses new risks, since the opportunity to tweak various parameters opens new potential biases in

how we view and interrogate data. With visual data in particular, image capture and analysis programmes have a range of options that make it easier to modify data presented, to change dimensions and alter contrast, to stitch together or merge multiple images, or to remove elements entirely. While such programme features are introduced for good reasons, they also create opportunity for beautification or worse still, manipulation of data with the intent to deceive.

Let us take Western blotting as an example. For many microbiologists, Western blotting will be a core skill used frequently in their day-to-day research, but for others will be something done only a few times in a year. Where traditionally the signal on a Western blot would have been detected using photographic film, often with multiple exposures of differing length, it has become increasingly commonplace to use charged-coupled device (CCD) camera technology to detect fluorescent signal from a blot. CCD cameras can have many advantages for developing Western blots in terms of ease of use, greater sensitivity and dynamic range, reduced background noise, generating digital data. However, the programmes they use also allow for various parameters to be altered that affect the way the blot appears, often before the data is even saved. This allows specific lanes, or groups of bands, to be selected and modified to enhance particular features of the data or decrease background signal and then saved as original data, without the context provided by the entire blot. In addition to changes in the way that blots are developed, long-established best practices such as inclusion of molecular weight markers, staining blots to check equal loading of each lane or consistent transfer from gel to membrane, and using suitable positive and negative controls, are increasingly falling by the wayside. The result is that for a significant number of the papers submitted to *Nature Microbiology*, Western blot data is of questionable quality and it can be challenging to distinguish cases where manipulation has taken place from untainted data that has simply not been generated and stored appropriately.

Like many other journals, in addition to asking authors to complete a reporting checklist to ensure that key experimental

approaches are adequately described, for Western data we require authors to provide original raw data for all blots and gels in articles to be accepted. We then editorially assess whether the data included in figures correlates with the source data and undertake integrity checks to look for signs of manipulation. We do not expect all blots to be immaculately presented; blots can be 'ugly' and still be good data. However, we strongly recommend that the raw data initially generated includes the entire blot (not just selected lanes), with molecular weight markers clearly labelled and suitable loading controls provided. Removal of certain lanes from a blot is acceptable if they are not pertinent to the scientific point being supported, but any splices should be clearly delineated in the figure and the relevant lanes noted on the raw data. Given the mobile nature of the research workforce, groups should also establish standard procedures for how their data is stored and catalogued, so that raw data for any given figure can be accessed and re-analysed at a future point, even after the individual that generated the data has moved on. In cases where we cannot be confident that the data is real, unmodified and matches the source data, or where the source data cannot be found, we will not proceed with publication.

Of course by the time we get a chance to see any data, if any corners have been cut it will have happened many months, if not years, previously. If a microbiologist is about to undertake an experiment for which they are not previously experienced, whether technically demanding or more straight-forward, we recommend seeking advice and input from researchers experienced in the technique (lab mates, collaborators, institutional colleagues or others in the field) at the experimental design stage to establish best practice. Proceeding with less haste in these early stages can actually end up saving time in the long run if it means that the data output of an experiment is well-controlled, appropriately described and informative. □

Published online: 23 January 2019  
<https://doi.org/10.1038/s41564-019-0365-1>