Current situation

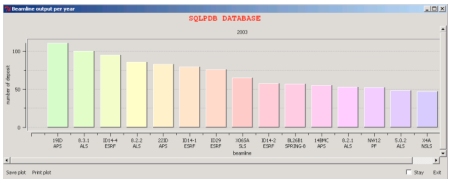
W0.308

Teaching Elves to Collect Data: An Analysis of the Last Million Diffraction Images from ALS 8.3.1.
James Holton, Physical Biosciences, Lawrence Berkeley National Laboratory, Berkeley, CA 94720.

Most X-ray data sets collected at synchrotron sources do not produce usable results. An analysis of data collected in 2003 at the ALS beamline 8.3.1 shows that 2346 datasets were collected and 41 structures were deposited in the PDB. Although it is understandable that not every dataset leads to a published structure, it is troubling that ~98% of them do not. This large gap between collected data and useful results is not unique to 8.3.1. The 28 operating American PX beamlines collect ~100,000 datasets/year. This suggests that a great deal of improvement in scientific productivity can be attained if the reasons for failed projects are better understood.

57 datasets/deposit !!!!

How productive is 8.3.1 ?



imgCIF

- Can we improve beamline productivity by imgCIF ?
- Is it possible to introduce uniform format or at least somewhat uniform header ?

Current situation
global view *vs.* local view

Company I

- ESRF
- NSLS
- ALS
- APS
- APS-ID19

Current situation
global view *vs.* local view

Company II

- 2 major frame formats
- Indefinite variability of the frame formats
  28 pages header documentation
- Random generator of space signs
- Free descriptions of goniostats

Image manipulation
by experimental system

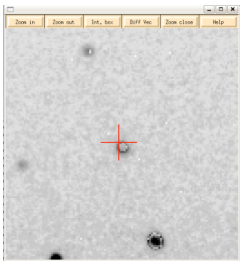


Image manipulation
by experimental system

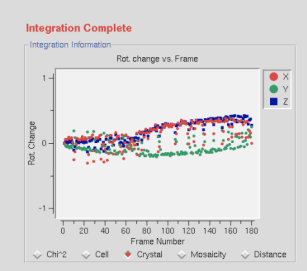
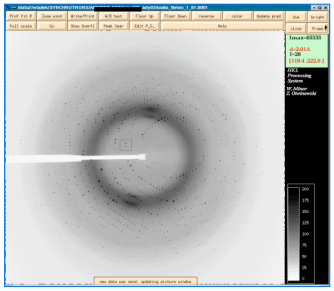
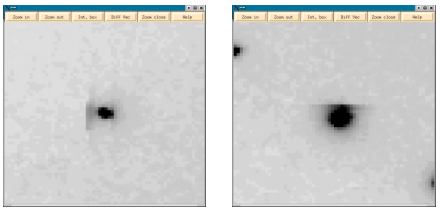


Image manipulation
by Company III



Effects for data



Rejections

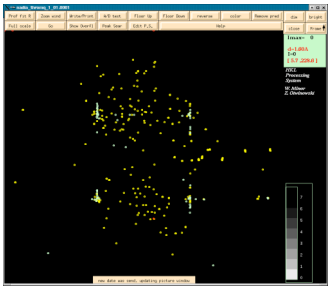


Image manipulations

Not picture-perfect

Nature's new guidelines for digital images encourage openness about the way data are manipulated.

Researchers struggle to amass good data and present them in as clear a fashion as possible. But what do we mean by 'clear' when it comes to images? In days gone by, whether we liked it or not, data acquired at the bench were not much different from what was published. In a biomedical lab, for example, samples that had been radio-labelled and separated on a gel were recorded on X-ray film. Composite figures were assembled, with lettering carefully placed around the mounted film. If a control was forgotten or a gel was uneven, the graduate student or postdoc was sent back into the lab to get it right 'for publication'. If a speck of dust on the film obscured data in the original photograph, another picture was taken. Slicing films to rearrange the order of samples, or to splice in a control group that was actually part of another gel, was not common because it took almost as much skill to do that as to rerun the experiment.

It is doubtful that scientists were more angelic then than now. It is

Image manipulations

more likely that, when it came to image manipulation, they wouldn't because they couldn't. These constraints led to the accepted standards for publishing quality images: what you got is what you saw. It's not that researchers didn't aspire to perfection — to have obtained images worthy of the admiration of colleagues enhanced your prestige because it proclaimed technical mastery. But the numerous examples of slightly inadequate data continually reminded everyone just how difficult it was not only to perform the perfect experiment, but to acquire the perfect image.

Digital image acquisition and processing tools have removed the physical impediments to perfect images and laid bare the inadequacies of current training practices. Traditions for image handling were not passed down from one generation to the next because there weren't any traditions. Into this vacuum has crept 'beautification' — the digital manipulation of properly acquired data for the purpose of making a figure cleaner, more perfect and more consistent with the best images yielded in such experiments. Removing dust from a digitized photo with the erasure tool, cropping bands from gels, and playing with fluorescence micrographs to enhance a particular effect are all attempts to show better results than were actually achieved in that run. In all these cases the data are legitimately acquired but then processed to yield an idealized image.

In *Nature*'s view, beautification is a form of misrepresentation. Slightly dirty images reflect the real world. Accordingly — and after consulting with technical experts — the *Nature* family of journals has developed a concise guide to appropriate image handling (www.nature.com/nature/authors/submissions/images), which will soon be incorporated into Guides for Authors.

In short, any digital technique that isn't applied to the entire image is suspect and needs to be explicated to the reader in the Methods or Supplementary Information. Authors should detail the instrument settings and the software manipulation that was performed on figures in an additional table in Supplementary Information. The fewer manipulations, the easier it will be for authors, referees and readers. Any changes made to misrepresent the data in the original image (such as boosting the contrast to eliminate backgrounds, making a collage of cells in a micrograph to show more cells in a visual field, or removing possibly informative bands in narrowly cropped gels) are of course strictly off limits.

> "In short, any digital technique that isn't applied to the entire image is suspect and needs to be explicated to the reader."

We should conspire to end the fetish of the perfect image. Let's all get a little more 'real'. *Nature* is happy to work with others to aid the promulgation of image standards that we can all live with. The responsibilities of the institutes that train students, of the investigators who use their labour, and of the journals that publish the data can be better defined. Finding ways to regain our trust in scientific images is a goal on which we can all agree. ■