Trasande, Caitlin, and Timo Hannay, "Changing the Conduct of Science: A Publisher's Perspective"

Changing the Conduct of Science

A publisher's perspective

Caitlin Trasande <c.trasande@us.nature.com>
and Timo Hannay <t.hannay@nature.com>

8 November 2010

Science is the ultimate collaborative, global human endeavour. The internet is the ultimate collaborative, global communication medium. They seem made for each other (and the web was, literally, made for scientists). Yet the whole-hearted adoption of these technologies to further the goals scientific research and accelerate the pace of discovery is neither natural nor inevitable. Numerous barriers slow and even halt that progress, some technical or practical, others social or psychological. This document attempts to identify some of these hurdles and briefly describe ways of overcoming them.

DATA ACCESS

Encourage the creation of good software tools. By and large, the foundational infrastructure that might enable scientists to organise, annotate and share their data already exists. True, data is accumulating at an awe-inspiring rate. But even if we cannot capture and process it all, we can in principle achieve a great deal with a lot of it because storage, bandwidth and computing capacity have never been cheaper or more abundant.

However, harnessing this power is not easy. The scientist himself is often the rate-limiting step for optimally processing data. For example, skill sets needed for managing digital data (e.g. sequences or images) vary dramatically depending upon field, institute and laboratory. Some scientists (particular those in the physical sciences) may be naturally adept and skilful digital data managers because working programmatically with data is the norm in their field, whereas scientists in other fields may lack any formal (or informal) training in managing digital data. It should be noted that while the average bench scientist does not need a data centre, almost every modern scientist needs software to analyse (and often generate/collect) data. The following might help their efforts:

- Foster the creation of metadata standards (like MIAME), as well as the expectation that scientists will routinely use them to annotate their data.
- Encourage the development of more and better software to make these tasks less time- and labour-intensive. (Ideally, data annotation ought to be a completely natural and integral part of the process of conducting experiments.)
- In particular, encourage the development of commercial software for researchers (for example by making it clear where government-funded providers will and won't operate, and by earmarking a certain proportion of grant funding for the purchase of software tools).

- Engage not just with publishing houses but also software houses and research-equipment suppliers.
- Give credit to those researchers who are genuinely open with their data, particularly when the data are used by others as a basis for their own research (see Attribution section below).
- Develop a consensus of what basic digital data management and computing skills are necessary to support high-quality data collection, processing, annotation and management.
- Fund training programs or the development of courses to ensure that all scientists have adequate competency in digital data management.

KNOWLEDGE ACCESS

Reward knowledge sharing of all kinds. Modern science was born when a reward structure was created that encouraged researchers to share their findings with each other through pages of academic journals rather then keeping them to themselves. Unfortunately these same incentives now have the perverse effect of discouraging knowledge sharing by other means. And as the opportunities for communication in the online world multiply (discussion forums, recommendations, wikis, file-sharing sites, blogs, microblogs, comments, votes, and so on), the aggregate cost of these lost opportunities grows. Whilst it is true that not all of these new means of communication are equally well suited to scholarship (and perhaps that some of them are downright counter-productive), the main reason that they are hardly exploited in research is the fact that contributions of these kinds are not tracked or rewarded. In the current incentive structure of science the author of the most influential academic blogs is trumped by the author of the most inconsequential peer-reviewed paper. This is patently wrong. Here are some ways it might be righted:

- Explicitly reward acts of self-arching in funder, institutional or other repositories by tracking this activity, making the statistics available, and using them in funding and appointment decisions.
- Similarly encourage the acts of posting preprints and blog entries, as well as commenting on them. (Systems for ranking these contributions by quality will be required, and any such system is vulnerable to gaming, but countermeasures are also possible so this is not an insurmountable challenge.)
- Conversely, recognise that in certain circumstances access restrictions are a feature, not a bug. (For example, where certain types of medical information are concerned, and where a truly open discussion can only take place away from the gaze of the globe and posterity.)
- Design a reward system for scientists who make themselves (and their reagents, algorithms, etc.) available to others. This could be piloted by tracking explicit acts of mentoring (e.g. evaluating PhD supervision).

ATTRIBUTION

Support and use ID systems for researchers. It is a truth universally acknowledged that actions are driven by incentives, and incentives by attribution and credit. A substantial change in the way that science is conducted is difficult to imagine without a corresponding change in the way that academic credit is tracked and assigned. It is unfortunate, therefore, that perhaps the only scientifically significant objects in the known universe that lack a robust identification system are scientists themselves. Until the assignment and use of personal identifiers becomes routine, it will be next to impossible to track and reward the wide range of activities in which a 21st-century scientist ought to be engaged. Here are some ways of making it happen:

- Support and use of identifiers for researchers as a way of assigning credit for a wide variety of contributions, and making the decisions that stem from this. Encouraging the use of ORCIDs would be a good start: publishers should make their creation and capture an integral part of the editorial process; funders should use them to reward contributions to the common good.
- While it is unrealistic (and arguably undesirable) to aim for One True Identity System, a wild proliferation of systems would be counter-productive, so the creation of new ones where existing ones suffice should be avoided. Furthermore, interoperability with other identity systems is key – any system that does not readily interoperate with others does not deserve support.
- Encourage researchers to see their IDs more like loyalty cards (i.e., a means to gain credit for their contributions) than as social-security numbers (i.e., oppressive instruments of a potentially intrusive bureaucracy). Instill confidence that identities and related data are secure. In this regard, useful lessons might be learned from certain consumer markets.
- Recognise that identity (e.g., ORCID) is different from authentication (e.g.. OpenID). Though the two are related, they are best kept distinct and should not be confused with one another.
- Encourage the wide dissemination of activity data associated with personal IDs, and hence the creation of a wide range of derived metrics and rankings. Critics who point out that scientific research is too complicated to be measured are correct, which is precisely why we need a proliferation of metrics to encapsulate this complexity. This can only be provided by an open, competitive market for metrics.

DATA GENERATION: PLACES, PEOPLE AND TRAINING

Support hubs of scientific activity and training. Core facilities are institutionally managed shared experimentation resources (e.g. DNA and protein sequencing, light and electron microscopy, mass spectrometry). They are professionally staffed and designed to provide expert-led access to speciality equipment and technologies. Core facility directors and staff often provide an array of services, including training on specimen or sample preparation, operation of equipment and software, data collection and analysis, as well as experimental design and interpretation of experimental results. Core facilities represent a unique physical space where scientists from different fields cross paths – and cross-fertilize ideas – in the course of carrying out their experiments.

Owing to their central and influential role in the creation of data and dissemination of specialized knowledge, **core facilities represent a valuable nerve centre of data-centric scientific activity.** As such, core facilities staff members are well positioned within the scientific network to propagate good data-related habits across an institute's research staff. To best make use of their role in science it would be valuable to:

- Engage professional societies (e.g. The Association of Biomolecular Research Facilities) in identifying and profiling core facilities at research institutes. Centrally maintain these profiles. Keep these facility and staff profiles up–to-date (e.g. as a condition of receiving ongoing federal funding).
- Identify key areas of in which good data-related habits would be most beneficial to the widest scientific audience. (For example, standardizing the annotation of experimental conditions in live cell imaging experiments.) Establish professional standards for each of the key areas.
- Create incentives for core facility staff (e.g. develop standards for crediting and attribution) to both a) provide the highest quality support to their communities, and b) disseminate locally developed knowledge across core facilities performing similar services.