The Open Science Proposal:
an outline of Open Science, with a section on the relevance of IT

(This is a slightly edited reprint of an article published in
SGR Newsletter, No 21, Autumn 2000, pp 8 - 9)

Abstract
The Open Science Proposal is that it would be useful for the scientific community to have a widely recognised Protocol of openness, applying to individual scientific projects. Only a small fraction of all scientific projects would be done according to the required standard of *radical openness*, but those projects could claim a 'gold standard' of openness. Such projects would permit detailed scrutiny and criticism, and so furnish an efficient route to socially established reliable knowledge. In the second part of the article I discuss some aspects of OS which relate particularly to Information Technology.

Open Science
In the Open Science Proposal, I suggest a specific modernisation of current ideas about openness in science. The starting point is the recognition that there are places in society for such contrary qualities as reserve, discretion, tact, and secrecy (Bok 1982). The next point to note is that the word science is nowadays freely used to cover a wide range of activities, and these activities are not merely *linked* with society, they are *embedded* in society. Given this situation, we need to step back from what Sissela Bok called the ritualistic denunciation of secrecy in science.

I use the following terms to describe four levels of openness in science. The first three already exist. The fourth may come to be, if we wish it ...

--- Secret Science: even the existence of the project is concealed

--- Restricted Science: publication of the results is subject to strict limitations in respect of timing and level of detail. Most commercial and applied government (including military) science is in this category

--- Circumspect Science: scientists publish when the project is complete, but till then are quite 'close'. Academic science, as practised to date, and when not Restricted, is in this category

--- Open Science: is the subject of this article. It has a precise meaning, namely the set of projects which are done according to the Open Science Protocol.

The Open Science Protocol is still at the embryonic stage of development. Its core is that the following stages of an Open Science project are all open, as they occur ...

- passage through institutions' ethics and safety committees

- application for funding

- review by funding body

- funding body's terms for support

- institution's terms

- log of the course of the project

- reports

- manuscripts submitted for formal publication

- referees' comments, revisions, published papers

- archiving of concise but detailed records of all the above stages.

Obviously, there are many aspects of this proposal which demand elaboration - more than is possible in this short article. Some aspects not treated here may be found in Cottey 1998. The questions which I have found to be raised most immediately by scientists centre around the belief that openness during a project is utopian ...

Why would any scientists go to the extra trouble and 'risk' of doing a project in a completely open way? Answer: there would be an important *incentive* - the enhanced credibility of scientific work done in a fully open way and withstanding the scrutiny of all who wish to scrutinise.

What of the risk of a smart competitor taking your promising idea and pipping you at the post of definitive proof? Answer: There is this risk, but most of us have an inflated conception of how interesting our latest bright ideas are to others. A much more common problem is to get others to take them seriously! And by posting one's progress, essentially in real time, one is putting down markers of possible priority much more efficiently than happens with the current (publish-when-everything-is-complete) practice.

Would not such a warts-and-all exposure be horribly embarrassing? Answer: Disclosure in a secretive milieu has a disproportionate impact -

Secrecy → Sensation! (on disclosure) → Fear of Openness → Secrecy.

By contrast, in the OS way, objections of a named referee, or disagreements on research in progress, would be no great deal - merely single items in a large amount of available information.

Open Science and IT

Doing a project the Open Science way will obviously involve a large amount of recording, storing and accessing of information. Is it

practicable? Before the IT era, the answer would be 'hardly'.

IT has transformed the situation. I see the Open Science way as taking its place alongside the three traditional ways of doing science (Secret, Restricted and Circumspect). The Open way will probably never become the principal way. There is, indeed, no need for that. A small fraction of all scientific projects done according to radical openness would have an influence beyond their small number. Consider, for example, those areas which are bedevilled by mistrust between experts (whose expertise may be debased by peer-group special interest) and outsiders (whose less-informed judgements may be faulty). One such area is the biological effects of ionising radiations. Some OS projects in this area might be specially helpful. Setting up and completing an OS project in such a tense area would, however, be difficult. The earliest OS projects - and none have been attempted to date - should be uncontroversial, simple and short.

So - the number of OS projects need not be very large. Nor need the volume of data collected and preserved be unmanageable. It will usually be possible to perform front-end reduction and parameterisation of raw data in ways that do not compromise the principle of openness. Naturally the procedures and algorithms used will be published. Analysis further down the line, which is subject to more assumptions, will be recorded and stored in more detail.

Today, it is relatively easy to record, store and access large volumes of data. The principal data problem for the OS Proposal is, I suggest, the variety of the data. Openness, for the OS Proposal, goes beyond simple technical data. Information about diverse conditions, such as the funding application, the conditions of employment, breakdowns, etc should all be recorded. Some exceptions to complete real-time openness are, however, appropriate. Confidential parts of staff CVs are an example.

Two IT developments currently underway may be expected to ameliorate the problem of recording, storing and accessing heterogeneous data. One is the development of world standards for dealing, in a unified manner, with data in varied formats - documents, databases, graphics, sound recordings, etc. The WWW Consortium's RDF (Resource Description Framework) is an example (Berners-Lee 1999).

The other specially relevant IT development (Johnston 1999) is LIMS (Laboratory Information Management Systems). LIMS is normally for the integrated management of a commercial company's range of scientific projects. A similar system adapted to OS should provide two services - a tool making it easy for OS project staff to conform to the OS Protocol; and a set of IT standards which will make it sufficiently easy for all to access and comment on an OS project.

The development and take-up of OS will be slow in comparison with the rapid further development of the Web. By the time the early versions of the OS way are tested, the Web should be a very convenient vehicle for recording, storing and accessing OS project data. Problems connected with variety of objects, authorised editing, authenticity and archiving are far from trivial, but they should be soluble. The key is to refrain from incorporating software that is more open-ended than necessary. OS users should be offered as much flexibility in their use of software as is needed, but not more. The OS Protocol should include a requirement that the data be placed on the Web in formats which have, at that time, become world standards (fully recognised in practice, and not simply in theory). That the early versions of OS could not, for example, support projects generating very large amounts of data would not be a significant defect.

IT has a bearing on the choice of OS infrastructure. One, centralised, model would require an Open Science Institute to keep a fairly tight rein

on OS practice. The more attractive alternative would be to follow the philosophy of the World Wide Web and the WWW Consortium. Successive versions of basic OS standards, principally the OS Protocol, would be published as a consensus by those interested. Then any investigators could unilaterally declare themselves as doing a project the OS way. There would be no formal, centralised monitoring but all outsiders could judge for themselves the quality of adherence to the OS way, and the quality of the work itself, including the quality of any open refereeing that had taken place. The build-up (or decay!) of acceptance of a project's claims would occur partly by open comment on the Web. OS would become a visible college.

A few words are needed about archiving ... The OS Protocol should include a requirement of arrangements to archive for a fixed period of time. Thirty years should be sufficient to permit most work to be absorbed into the canon of reliable scientific knowledge, or rejected, or quietly forgotten. In principle, an archive review would also take place after 30 years, so that projects of continuing interest could be preserved longer.

Technology drift is sometimes mentioned as a problem with archiving - if data are recorded using hardware or software which becomes obsolete, they may become effectively unreadable or lost. I suggest that was mainly a problem of the early years of IT. Although IT continues to develop rapidly, there is today more awareness of archivists' needs. Now, if the advice of professional archivists is followed, OS records should be readily accessed for 30 years, with the option, if required, for importing the records into a new storage system at the end of that period.

References

Berners-Lee, T (1999) *Weaving the Web*, Orion

Bok, S (1982) *Secrets: on the Ethics of Concealment and Revelation*, Pantheon

Cottey, A (1998) *Open Science and Practicable Politics* (pp 88 - 96 of *Green Politics in Grey Times* - Ecopolitics XI Conference Proceedings, ed C Star, pub University of Melbourne)
Johnston, H (1999) *Scientific Computing World* issue 48, August/September pp 17 - 18

Alan Cottey, University of East Anglia. a.cottey@uea.ac.uk
The basic version of this page was created in 2000
and the page was last modified on 24 October 2013.