

Open Research Data

Submitted on 05 Nov 2014 by Martin Alvarez-Espinar

'Nullius in verba' ('on the word of no one' or 'take nobody's word for it') is the motto of the [Royal Society](#) - one of the oldest scientific bodies in the world.

A central tenet of the scientific method is that one scientist can repeat and thereby verify or refute another's work.

Although research documents (i.e., scientific papers, posters, etc.) are published and spread through specialised magazines and congress proceedings, **raw data and methodology used for those researches is not often published.**

This idea is a perfect match for the aims of open data and there is an enormous amount of effort going into making the data available. In the USA, Phil Bourne has been very active in his [role as Associate Director for Data Science](#) at the National Institute of Health. The Australian and US governments, together with the European Commission, have funded the establishment of the [Research Data Alliance](#) - a global body to establish standards and best practice for the sharing of research data. [DatCite](#) is an organisation dedicated to making it easy for researchers to cite each other's data.

A key issue in this regard, and an element that is currently missing and needs to be fixed fast, is that researchers do not always receive credit for citations of their data in the same way they do for citations of their papers. For any researcher, it is papers published and citations of those papers that matter, so the culture must change to give this credit. If researcher *A* uses a dataset generated and published by researcher *B*, doesn't *B* deserve some of the credit for *A*'s discoveries and insights?

This is recognised in the [joint statement by the US Committee on Coherence at Scale and the UK Open Research Data Forum](#) that lists **6 principles** of open data:

1. The **data that provides the evidence** for the concepts in a published paper or its equivalent, together with the **relevant metadata and computer code must be concurrently available** for scrutiny and consistent with the criteria of "intelligent openness". Data must be: discoverable, accessible, intelligible, assessable (e.g. the provenance and reliability of data), and re-useable.
2. The data generated by **publicly - or charitably-funded research** that is not used as evidence for a published scientific concept should also be made intelligently open after a pre-specified period in which originators have exclusive access.
3. Those who re-use data but were not their originators must formally **acknowledge** their originators.
4. The **cost of creating intelligently open data from a research project is an intrinsic part** of the cost of research, and should not be considered as an optional extra.
5. There are justifiable limits to this "intelligent openness" similar to exceptions for open data in general (personal data, security, etc.).
6. Existing processes, reward **structures and norms** of behaviour that inhibit or prevent data sharing or new forms of open collaboration **should be reformed** so that data sharing and collaboration are encouraged, facilitated and rewarded.

These Open Research Data principles bring **new opportunities of collaboration to make scientific knowledge grow and spread**: combination of data will produce new knowledge, and results

reported in scientific articles would be verified by third parties.

Anyway, releasing these resources is not trivial, and some challenges will always be present: **embargo on access to data until the results has been published** to avoid stealing of data by others; **citation of data** is not credited to researchers the same way as citation of scientific articles thereby limiting reward to researchers; **persistent hosting and curation** of research data is necessary to ensure long-term availability —of course, cost must be covered by someone (funding agencies, universities) to maintain free access.

Academia seems to be far from the traditional open data publishers. In fact of that, the amended [Directive on PSI re-use](#) does not apply to academic institutions, such as educational and research bodies, including intermediate organisations aiming at transferring of research results. Anyway, everyday there are more organisations pushing for this new paradigm in the scientific sector.

[This blog post has been written by the ePSI Platform Advisory Board members: Makx Dekkers, Phil Archer and Martin Alvarez-Espinar.](#)