Perspectives on Open Science and scientific data sharing: an interdisciplinary workshop

Authors : Giovanni Destro Bisol, Paolo Anagnostou, Marco Capocasa, Silvia Bencivelli, Andrea Cerroni, Jorge Contreras, Neela Enke, Bernardino Fantini, Pietro Greco, Catherine Heeney, Daniela Luzi, Paolo Manghi, Deborah Mascalzoni, Jennifer C. Molloy, Fabio Parenti, Jelte M. Wicherts & Geoffrey Boulton

Summary

The review article resumes an interdisciplinary workshop, “Scientific data sharing” organized by the Istituto Italiano di Antropologia in Anagni, Italy, 2-4 September 2013. It’s principal objective was to look at Open Science and Open Data from a broad perspective.

Workgroups dealt with four questions:

- Setting a common framework about open data principles, values, and opportunities;
- Insights into different scientific practices with OD in biology, psychology, epidemiology and archaeology;
- A case study of human genomics;
- Open science and the public.

Summary

Three proposals for future action emerged from the workshops :

- Integrate top-down initiatives by governments, institutions and journals with bottom-up approaches from the scientific community;
- Popularize the societal benefits of open science;
- Introduce arguments from social sciences and humanities in the educational dissemination of open data.

The concept of Open Science

The authors point out that the Open Science movement is growing, and is defined as making the publication of scientific concepts and their datasets easily accessible to all.

The movement touches “hard” and “soft” sciences, as well as government data.

From their point of view, Open Science is interdisciplinary but infrastructural, economic and motivational barriers remain obstacles to “cross-fertilization”.

Setting a common framework

If knowledge is one driving force in socio-economic exchanges, it is also a driver of inequality through the quality and quantity of information available and the possession of “cognitive capital”.

The dissemination and ancrage of scientific knowledge within society is a fundamental democratic bedstone aiding individuals to connect themselves to larger local and global dynamics.

Setting a common framework

What is a “knowledge society”?

The authors establish four phases aligning three stakeholders:

- Knowledge generation : Individual → Collective;
- Knowledge institutionalization : Collective → Knowledge;
- Knowledge diffusion : Knowledge → Collective;
- Knowledge socialization : Collective → Individual;

“Innovation” is the combination of these four phases. Science must be open to society and vice-versa.

Setting a common framework

The recent “Data storm” has undermined open publication of science concepts that were a key element in the scientific and industrial revolutions of the 18th and 19th century.

Data is no longer published leading to results that are not reproducible and undermine credibility in the scientific process. The authors consider this to be “scientific malpractice”.

Setting a common framework

Open data and data sharing offers opportunities for science to face new global challenges. It should be a public rather than private enterprise.

Open science requires “intelligent openness” of data and metadata and should be:

- Discoverable;
- Accessible;
- Intelligible;
- Assessable;
- Re-usable.

Setting a common framework

Three legitimate boundaries to openness:

- Commercial activities where there is an overriding public interest in deriving economic benefit;
- Personal data balanced with public interest;
- Safety and security, both physical and political.

Who decides where to set the limits? Scientists, universities, funders, academic publishers, learned societies, governments all have their role to play.

Setting a common framework

Two overlapping benchmarks:

- Open Knowledge Definition (OKD): “A piece of data or content is open if anyone is free to use, reuse and redistribute it”.
- The Panton Principles for Open Data in Science : all scientific data is clearly placed in the public domain via a copyright waiver or appropriate license..

Insights into scientific practices

What and how much data is shared? Why not?

The authors consider the impact of open science is “greatest when it operates as a collaborative process”.

Transparency and collaboration are binded intimately in order to maintain a continuous redefinition of theories based on the production of new knowledge, an essential condition for science.

Cooperation also optimises human and technical resources.

Insights into scientific practices

Sectoral empirical studies on practices in genetics and biodiversity reveal wide discrepancies in data sharing but common causes.

- Choosing which data is shared and maintaining control afterwards;
- Editing and reformatting data is costly and goes unrecognized;
- Personalized metadata formats impede sharing;
- Neither data management plan nor manager within their institutions;
- But willingness to share given policy and support.

Insights into scientific practices

The authors then present two counter-examples where data sharing plays a decisive role.

Insights into scientific practices

The SARS (Severe Acute Respiratory Syndrome) in 2003 caused 623 deaths for 7761 infections over three continents but was defeated in four months, thanks to international real-time data sharing and cooperation, coordinated by the World Health Organization.

Insights into scientific practices

Timeline :

- February 14: Outbreak in China;
- March 21: First clinical description of SARS;
- March 26: Virtual roundtable bringing together 80 clinicians from 13 countries and establishing a coordination of 11 labs in 9 countries to research on the aetiology of SARS;
- April 16: Identification of the pathogen, a member of the coronavirus family;
- May 1: Publication of the complete genome;
- July 5: Coordinated public health measures break the last human chain of transmission signaling the end of the pandemy.

Insights into scientific practices

An example from archeology reveals a certain number of paradoxes.

The Italian national database PreBiblio references 15000 prehistoric sites with bibliographical and geographical coordinate data. If the sharing this data is important for site management, research and conservation, free access could entail risks :

- Inviting looting and pillage;
- Public access to hitherto unknown sites;
- Misuse of sites because of tourism.

Insights into scientific practices

The European Commission has taken action in favor of Open Data.

Researchers must provide access within six months to their data used for articles coming out of EC funded studies. (Guidelines for Open Access, 2007).

At the same time, the OpenAirePlus project provides a networking and technical platform to assist national repositories in implementing Open Data.

The example of human genomics

The study of human genomics is perhaps the first and biggest project involving international data sharing between research groups.

The difficulties encountered help structure public policy in order to insure best use of the data.

The example of human genomics

Policymakers have to reconcile conflicting priorities among stakeholders :

- Open data should remain open;
- Scientists need to be recognized for their work;
- Biobanks have high data collection and management costs;
- Human subjects need protection against unwanted or unauthorised use of their data.

The example of human genomics

The American “Human Genome Project” evolved in 1996 to short-circuit concern about private patenting of genomes by imposing a 24 hour deadline for public submission of genome sequence sets.

To satisfy scientists anxious about use and credit for their data, the Fort Lauderdale Principles (2003) reaffirmed their rights. A moratorium of 6-12 months on publications using data from previous research gives a head start to the original scientists.

The example of human genomics

The nature of open data conflicts with processes that protect privacy. “Liquid surveillance” describes the transfer of state prerogatives towards a multitude of actors, a question still unresolved.

Biobanks must resolve an economic equation but also the contractual relations with human subjects where open data could mean loss of control.

Public institutions (CHRIS) point the way, where a mechanism of continuous information and exchange led to acceptance by 97% of donors and fostered their trust.

Opening science to society

Widespread diffusion of science through the internet gives access to the best (and the worst) research.

Science communicators compete with many other actors, blogs or social networks.

The scientific community needs to adapt rather than hide from the problem.

The authors call for an interdisciplinary debate on Open Science to more fully engage the public with science.

Conclusion

Further interdisciplinary initiatives need to address :

- Governments as major stakeholders;
- Allow for bottom-up creativity;
- Involve the public in the discourse about open data;
- Create awareness among students and researchers about Open Data;
- Examine the dichotomy about sharing or withholding data.

In conclusion, interdisciplinarity is the most promising avenue for advancing Open Science.