

## 2.5 Opening Science to Society: an interdisciplinary initiative for data sharing

Marco Capocasa<sup>1,3</sup>, Paolo Anagnostou<sup>1,2</sup>, Nicola Milia<sup>4</sup>, Cinzia Battaglia<sup>2</sup>, Valentina Coia<sup>5</sup>, Maria Enrica Danubio<sup>1,6</sup>, Fabrizio Rufo<sup>1,3</sup>, Emanuele Sanna<sup>4</sup>, Giovanni Destro Bisol<sup>1,2</sup>

<sup>1</sup>Istituto Italiano di Antropologia, Roma, Italia

<sup>2</sup>Dipartimento di Biologia Ambientale, Sapienza Università di Roma, Italia

<sup>3</sup>Dipartimento di Biologia e Biotecnologie "Charles Darwin", Sapienza Università di Roma, Italia

<sup>4</sup>Università di Cagliari, Dipartimento di Scienze della Vita e dell'Ambiente, Italia

<sup>5</sup>European Academy of Bolzano (EURAC-research), Istituto per le mummie e l'IceMan, Bolzano, Italia

<sup>6</sup>Dipartimento di Medicina clinica, sanità pubblica, scienze della vita e dell'ambiente, Università di L'Aquila, L'Aquila, Italia

The advent of the digital era is radically changing the way in which scientific results are produced and disseminated. However, to what extent this ongoing revolution might foster the advancement of science depends on our ability to make information available to the whole scientific community. It is only in the last two decades that the importance of a robust, effective and sustainable data sharing has been fully recognized by the scientific community (KING 2011, LANG 2011). Some important funding bodies are also moving in this direction. This is the case of the European Commission, which is launching a pilot project for the Open Data, in order to promote the sharing of scientific data that will be produced in the context of *Horizon 2020*, the new European program for funding research and innovation.

A widespread data sharing has numerous advantages for research progress. Firstly, the availability of large datasets increases the range of analyses that can be applied and the accuracy of the results and of the related implications. Secondly, data sharing is an essential requirement towards the definition of a transparent science. In fact, data access is a necessary step in order to verify the reproducibility of the scientific results and is a strong deterrent to falsification (FISCHER, ZYGMOND 2010). Finally, the possibility to build informative datasets using previously published data represents a non trivial benefit in economic terms. Making data fully available minimizes risk of research duplication, with an evident resource

saving. Moreover, data sharing is accompanied by a diversification of their use, even by researchers with innovative ideas but without the necessary financial support to put them into practice.

These positive aspects can help accelerate scientific progress. Their effectiveness depends on the implementation of efficient sharing policies by the founding agencies and the scientific journals. However, a robust and unconditional data sharing is not always easy to put into practice and the difficulties may vary depending on the research field. Some researchers have referred to the high economic costs for establishing the necessary infrastructure for data sharing. Ethical issues were also highlighted. This is particularly true for studies based on data related to human subjects, where risk of privacy violation and data misuse complicates the collection of new samples and the subsequent production and analysis of data (MURDOCH, CAULFIELD 2009; NELSON 2009; GIFFELS 2010; TENOPIR *et alii* 2011).

The increasing interest of the scientific community on data sharing is witnessed by the proliferation of publications and projects concerning the relationship between the amount of available data and the scientific progress. However, other points of interest are beginning to appear, particularly regarding the awareness on the importance to spread the "sharing culture" among young researchers and to attract on this issue the attention of the general public. With the aim to meet this wider view of the problem, we are launching the initiative *Opening Science to Society* by which we will deal in an integrated manner the scientific, educational and public aspects of data sharing. In the first stage of the initiative we want to focus our efforts on the analysis of "sharing behaviour" in human genetics, continuing the work we have already carried out (MILIA *et alii* 2012; CONGIU *et alii* 2012; ANAGNOSTOU *et alii* 2013). In the long term, our goal is to build a general framework which may be used in different fields of knowledge and help interdisciplinary interactions.

The project's three lines of action are based on the following different approaches in order to better understand the potential and the possible implications of data sharing:

1. *You can't manage what you don't measure*: the analysis of the degree and methods of data sharing using an ad hoc developed procedure (see MILIA *et alii* 2012) and the awareness of authors on open data, asking them to fill in a questionnaire. This action line is expected to have the following outcomes: (i) empirical evaluation of the degree of true data sharing (ii) evaluation of possible influences of the sharing policies of scientific journal and funding agencies on the degree of data release; (iii) identification of barriers (ethical, legal, professional) to the sharing of scientific data via administration of a questionnaire to the authors.

2. *Educating for the future*: the main objective of this line of action is to develop new educational tools to promote greater awareness on data sharing among university students and young researchers. The star-

ting point will be the conduction of a questionnaire-based survey aimed at understanding what they know and what they think about the topic of data sharing and its implications. The results will allow the definition of innovative proposals to implement data sharing in the teaching modules and thus to promote a greater culture of sharing.

3. *Bridging science to society*: we would like to promote interdisciplinary initiatives (e.g. consensus conference, public debate, on-line forum) to favour dialogue among scientists, policy makers and the general public in order to stimulate critical discussions on data sharing. This line of action represents an important meeting point for the definition and the implementation of guidelines by which the general public and "experts" can interact effectively, exchanging and integrating information, resources and knowledge.

We are aware that these objectives can be achieved only by sharing skills and experiences, ideas and opinions. This is the first challenge: the development of a collaboration between experts and scholars of different disciplines, united by the common good of a science open to society.

Our first move in these directions is the launch of the *Opening Science to Society* web site (<https://sites.google.com/site/openingsciencetosociety/>), a workspace we would like to share with all those who believe that the philosophy of open data is an important means to advance scientific progress and open up science to society. At present, the web site gives access to: (1) a brief synopsis of the initiative; (2) information about our ongoing activities; (3) a forum for discussion of scientific, educational and ethical aspects; (4) an updated list of articles concerning data sharing; (5) numerous links to scientific and educational resources.

We are very interested in collecting ideas and suggestions and to expand the cooperation. To do this, we invite you to contribute to the debate through our online forum (<https://sites.google.com/site/openingsciencetosociety/Forum>).

### Acknowledgements

This work is supported by the Istituto Italiano di Antropologia (Project: *Scientific data sharing in the digital era: a survey on genetic variation in human populations*).

### Bibliography

- ANAGNOSTOU P., CAPOCASA M., MILIA N., DESTRO BISOL G. 2013, *Research data sharing: lessons from forensic genetics*, in «Forensic Science International: Genetics», submitted.
- CONGIU A., ANAGNOSTOU P., MILIA N., CAPOCASA M., MONTINARO F., DESTRO BISOL G. 2012, *Online databases for mtDNA and Y chromosome polymorphisms in human populations*, in «Journal of Anthropological Sciences», 90, pp. 201-215.
- FISHER B.A., ZIGMOND M.J. 2010, *The essential nature of sharing in science*, in «Science and Engineering Ethics», 16, pp. 783-799.
- GIFFELS J. 2010, *Sharing data is a shared responsibility. Commentary on: The essential nature of sharing in science*, in «Science and Engineering Ethics», 16, pp. 801-803.
- KING G. 2011, *Ensuring the data-rich future of the social sciences*, in «Science», 331, pp. 719-721.
- LSNG T. 2011, *Advancing global health research through digital technology and sharing data*, «Science», 331, pp. 714-717.
- MURDOCH C., CAULFIELD T. 2009, *Commercialization, patenting and genomics: researcher perspectives*, in «Genome Medicine», 1, p. 22.
- MILIA N., CONGIU A., ANAGNOSTOU P., MONTINARO F., CAPOCASA M., SANNA E., DESTRO BISOL G. 2012, *Mines, yours, ours? Sharing data on human genetic variation*, in «PLoS One», 7(6): e37552.
- NELSON B. 2009, *Data sharing: Empty archives*, in «Nature», 461, pp. 160-163.
- TENOPIR C., ALLARD S., DOUGLASS K., AYDINOGLU A.U., WU L., READ E., MANOFF M., FRAME M. 2011, *Data sharing by scientists: practices and perceptions*, in «PLoS One», 6(6):e21101.