EDITORIAL

Prof. Dr. Luiz Olavo Bonino da Silva Santos
GO FAIR International Support and Coordination Office, Leiden University Medical Center. E-mail.:luiz.bonino@go-fair.org.

THE ATTENTION TO THE APPROPRIATE DATA TREATMENT IS FUNDAMENTAL TO POSSIBLE ACHIEVE ALL THE POTENTIAL TO THE OPEN SCIENCE

Since XVII century, the societies demand a greater opening and transparency for the academic activities. Considering that most of scientific founding come from government resources, it is natural that the society expects a better access to the elements that constitutes the scientific research, like data, experiments and results. This causes not only the society access to those results, but also that other scientists may reuse data and results of other researches in another different way and reach other conclusions. Therefore, a greater data reuse and results from a more opened scientific environment enables a more efficient and effective use of the resources invested.

In the last decades, the development of computer technologies allowed a large growth of the communication capacity, and, therefore, also the accelerations of the data production and availability. On the one hand, all this increase in the volume of available data brings clear benefits. For example, when we have more data about a subject, we can make more informed and precise decisions. We have the possibility to answer complex questions because we have a greater variety of information and data. On the other hand, the increase of data volumes, variety and complexity makes more difficult the understanding of those data collection constituent elements and the selection of those which are more relevant to our problems.

The data management is becoming a critical issue in the organizations that use and/or produce them. In the Universities, we have a great diversity of colleges and researches groups within these colleges that collaborate on an even larger number of projects with other partners. Those projects frequently produce, consume, transform and analyze data of various types and sources. Information about how the experiments are projected and conducted is crucial to guarantee the reproducibility of the research. That is, the research results can be systematically reproduced and validated.

With the object to discuss solutions for those problems related to proper management of research data, a group of scientists, leaders of development and research organizations and data management experts met during 4 days for a workshop called Jointly Designing a Data FAIRPORT1, in Lorentz Center, Leiden, Holland, in January 2014. The discussions in that workshop, it has emerged the FAIR data movement, promoting activities for data and other elements of the academic activity to be findable, accessible, interoperable and reusable. Since then, the FAIR initiative has gained wide international recognition, especially after March 2016, The FAIR guiding principles for scientific data management and stewardship was published in the Scientific Data Journal, of Nature. This paper presents a set of 10 principles that aim to guide data management activities to set those data increasingly available, accessible, interoperable and reusable.

The FAIR principles are clearly nothing new. They are only a convergence of decades-long concerns in the scientific community about how to better manage their artifacts. This convergence and the current moment of great concern for efficiency and return on investment has made FAIR principles reach the highest organizational and political spheres. For example, in 2016 the FAIR principles were officially endorsed by the G20 group of countries mentioned in item 12 of the official report of the G203 meeting in China. In 2017 the same happened at the meeting of the G74 group of countries in Turin, Italy. In Europe, FAIR principles have been adopted by the European Commission and since January 2017 have been part of the requirements for the submission of Horizon 2020 research project proposals as well as forming the basis of the European Open Science Cloud (EOSC) the European open science support infrastructure. Development agencies in the US and other parts of the world have also adopted FAIR principles as essential requirements in their funding programs.

In Brazil, as it should be, data management is also a growing concern in research institutions. Universities, research centers, development agencies and other associated organizations have been in constant debate about which approaches, techniques and technologies should be adopted at the organizational, regional and national levels so that the country can observe a better organization and consequent improvement in activities scientific research.

Everybody happens a similar debate and one of the common elements is the perception that the problem will only be solved with a multidisciplinary approach where the researchers form synergetic partnerships with professionals of other formations. This diverse set of professionals brings together the scientist, who knows the domain where data is created and / or reused, the library professional, whose training provides the necessary knowledge to organize collections of information, and professionals in the technological area, such as software engineers, ontology engineers, data specialists, etc., who create, adjust or operate the technological tools necessary for such data management. These multidisciplinary teams should work in a harmonious way, dividing the tasks according to their areas of responsibility, and having as a common goal a better management and processing of the data so that the scientific process is more effective and efficient.

The current world trend is the adoption of this multidisciplinary organization for data management based on the FAIR principles that provide the fundamental guidelines for this management. An example is the accessibility aspect of the FAIR principles, which does not require data to be opened, but that the access protocol is well defined, open (the protocol, not necessarily the data) and widely implementable. This makes sense in the biomedical field where diverse types of data such as human patient data cannot be opened, including due to legal issues related to security and privacy. Obviously not open-ended science, data should be preferably open, but, as already exemplified, this is not always possible. However, even if they do place restrictions on the accessibility of some data, these restrictions and the necessary conditions for access to them must be explicit and well-defined so that researchers or any other potential users of such data can be clear about what conditions and what procedures should be followed so that they can be used in scientific investigations.

Brazil can integrate, from the outset, this global movement and actively participate in the definition of the guidelines, rules, technologies and approaches that will be used worldwide. The country has relevant capacity in the areas required for FAIR data management and there are already several national initiatives for these principles to be adopted on a larger scale when we can then take advantage of the benefits that the approach offers us. We must work from now on to coordinate these activities to avoid large divergences in implementation, redundancy and unnecessary rework, and broaden participation by drawing more people and organizations into this movement.

References

1. Jointly designing a data FAIRPORT, Lorentz Center, Jan 2014. https://www.lorentzcenter.nl/lc/web/2014/602/info.php3?wsid=602.
2. Wilkinson, M.D., et al, The FAIR guiding principles for scientific data management and stewardship, 03-2018, Scientific Data, https://doi.org/10.1038/sdata.2016.18.
3. G20 Leaders' Communique Hangzhou Summit, Set. 2016, Hangzhou, China, http://www.g20chn.org/English/Documents/Current/201609/t20160906_3395.html.
4. G7 Science Ministers Communiqué, Set. 2017, Turim, Italia, http://www.g8.utoronto.ca/science/2017-science-communique.html.